



University  
of Glasgow

Rousselet, G.A., Gaspar, C.M., Wieczorek, K.P. and Pernet, C.R. (2011) *Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects)*. Frontiers in Psychology . ISSN 1664-1078

<http://eprints.gla.ac.uk/56171/>

Deposited on: 27 September 2011



# Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects)

Guillaume A. Rousselet<sup>1\*</sup>, Carl M. Gaspar<sup>1</sup>, Kacper P. Wierzchorek<sup>1</sup> and Cyril R. Pernet<sup>2</sup>

<sup>1</sup> Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

<sup>2</sup> Brain Research Imaging Centre, SINAPSE Collaboration, University of Edinburgh, Edinburgh, UK

## Edited by:

Paul Sajda, Columbia University, USA

## Reviewed by:

Marios G. Philiastides, University of Nottingham, UK

Michael Tangermann, Technische Universität, Berlin

## \*Correspondence:

Guillaume A. Rousselet, Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, College of Medical, Veterinary and Life Sciences, 58 Hillhead street, Glasgow, G12 8QB, UK.  
e-mail: guillaume.rousselet@glasgow.ac.uk

We studied how task constraints modulate the relationship between single-trial event-related potentials (ERPs) and image noise. Thirteen subjects performed two interleaved tasks: on different blocks, they saw the same stimuli, but they discriminated either between two faces or between two colors. Stimuli were two pictures of red or green faces that contained from 10 to 80% of phase noise, with 10% increments. Behavioral accuracy followed a noise dependent sigmoid in the identity task but was high and independent of noise level in the color task. EEG data recorded concurrently were analyzed using a single-trial ANCOVA: we assessed how changes in task constraints modulated ERP noise sensitivity while regressing out the main ERP differences due to identity, color, and task. Single-trial ERP sensitivity to image phase noise started at about 95–110 ms post-stimulus onset. Group analyses showed a significant reduction in noise sensitivity in the color task compared to the identity task from about 140 ms to 300 ms post-stimulus onset. However, statistical analyses in every subject revealed different results: significant task modulation occurred in 8/13 subjects, one showing an increase and seven showing a decrease in noise sensitivity in the color task. Onsets and durations of effects also differed between group and single-trial analyses: at any time point only a maximum of four subjects (31%) showed results consistent with group analyses. We provide detailed results for all 13 subjects, including a shift function analysis that revealed asymmetric task modulations of single-trial ERP distributions. We conclude that, during face processing, bottom-up sensitivity to phase noise can be modulated by top-down task constraints, in a broad window around the P2, at least in some subjects.

**Keywords:** ERP, GLM, single-trial analyses, bootstrap, N170, individual differences, faces, noise

## INTRODUCTION

Following the first reports of larger scalp responses to faces compared to objects (Bötzel and Grüsser, 1989; Jeffreys, 1989; Jeffreys and Tukmachi, 1992; Jeffreys et al., 1992; Seeck and Grüsser, 1992), there have been hundreds of studies on the early event-related potentials (ERPs) to faces and objects. The vast majority of these studies used (i) averaged ERP, (ii) group statistics, and (iii) categorical designs. Their findings can be summarized shortly: sometime between 100 and 200 ms after stimulus onset, ERPs to different object categories tend to differ from each other, and faces are most of the time associated with larger N170 peaks than other object categories (Rossion and Jacques, 2008).

Recently, several research groups have started to study these early preferential responses to faces in individual subjects (Schyns et al., 2003, 2007; Philiastides and Sajda, 2006; Philiastides et al., 2006; Rousselet et al., 2007a, 2008a,b, 2009, 2010; Smith et al., 2007; Liu et al., 2009; Ratcliff et al., 2009; van Rijsbergen and Schyns, 2009). Individual subjects' ERPs, show, not surprisingly, systematic differences between faces and objects consistent with group effects reported so far (Rousselet et al., 2008a). These studies have also revealed inter-subject differences: despite coarse agreement between group and individual subject statistical analyses, individual

subjects show reliable ERPs to faces and noise, which differ reliably across subjects, for reasons yet to be discovered (Rousselet et al., 2010; Gaspar et al., in press Reliability of ERP and single-trial analyses).

Individual differences in early visual processing have been largely ignored in the face literature, and implicitly treated as measurement errors that can be filtered out by averaging data across subjects. Although understanding the average brain is a worthy goal, only the single-trial approach, in conjunction with parametric designs, will allow us to understand brain mechanisms and the information content of brain states (Schyns et al., 2009; Schyns, 2010). In the single-trial framework, timing is essential. Indeed, how fast the visual system can discriminate among object categories provides strong constraints on possible computational implementations (Thorpe and Fabre-Thorpe, 2001; Rousselet et al., 2004; Thorpe, 2009). In particular, the timing of task modulations might help us tease apart periods of mostly bottom-up, stimulus driven activity, from time-windows engaging flexible neuronal populations that might be tuned to certain tasks. Thus, task modulations are key to understand brain mechanisms (Schyns, 1998; Pernet et al., 2007; Schyns et al., 2009).

For over 10 years, the ERP face literature has been debating the existence of task modulations of the N170 face preferential response. Several studies used targets vs. non-targets manipulations, in which faces at fixation are attended or ignored, and found no evidence of task modulations on the N170 (Séverac-Cauquil et al., 2000; Carmel and Bentin, 2002; Rousselet et al., 2007b), and its magnetic analog, the M170 (Lueschow et al., 2004; Furey et al., 2006; Okazaki et al., 2008). Similarly, intracranial recordings failed to reveal top-down modulation of the N200 to faces (Puce et al., 1999). One exception is found in a recent study, which reported an effect of category expectation on the N170 (Aranda et al., 2010). However, the effect seems to be weak and in a direction opposite to the one expected, so it might be a type I error. In contrast to target vs. non-target task manipulations, the N170 can be modulated by spatial attention (Jacques and Rossion, 2008; Crist et al., 2008) or by directing attention away from faces, in conditions in which letters superimposed on faces have to be discriminated (Eimer, 2000; Mohamed et al., 2009). Effects of language interference (Landau et al., 2010) and working memory (Sreenivasan et al., 2007) have also been suggested.

Hence, at least in some conditions, early face processing seems to be modulated by spatial attention and other factors. However, the modulations of face ERPs reported so far tend to be ill defined because it is unclear what aspect of face processing is modulated by the task. It remains also unclear how and when task demands affect the processing of a face presented at fixation. Very few studies have tackled this fundamental question by using a design in which the same stimulus is presented but processed differently because task requirements change the diagnosticity of input information (Pernet et al., 2007; Schyns, 1998). Schyns and his colleagues used reverse correlation techniques and large number of trials to reveal changes in single-trial information sensitivity (Schyns et al., 2007; Smith et al., 2007; van Rijsbergen and Schyns, 2009). However, although these studies show that ERPs are sensitive to different information from the same stimuli in different tasks, they do not provide a quantification of how task requirements affect the brain sensitivity to the same information.

One earlier study aimed at answering this question and reported larger N170 amplitude in a gender task compared to an identification task, but only for coarse, not fine scale information (Goffaux et al., 2003). This result suggests the use of certain face spatial scale information when it is relevant for the task at hand, an interaction between task demands and available information essential to reveal the information content of brain activity (Pernet et al., 2007; Schyns et al., 2009). However, the effects reported by Goffaux et al. (2003) seem very small and there was no report of task effects in individual subjects. It is also unclear if the effects might not be due to task modulations of the ERP sensitivity to the structured noise added to filtered images.

A more recent study reported one of the most striking and interpretable task effects on ERP face sensitivity. In Philiastides et al. (2006), experiment 2, a cue indicated on each trial how subjects were to process a subsequent stimulus: either discriminate its color (red vs. green) or its category (face vs. car). Colored pictures of faces and cars had two noise levels, created by altering the Fourier phase spectrum, which contains most of the information about object identity (Gaspar and Rousselet, 2009). In the categorization task,

single-trial activity discriminated between face and car trials. In the same task, ERPs to faces and cars were sensitive to the level of image phase noise roughly in the time period 100–300 ms after stimulus onset. Among several important results in this paper, the authors show that in the color task, in which the noise dimension becomes task irrelevant, noise sensitivity was strongly reduced shortly after 200 ms. This is an important result because it suggests a timing for task effects in situations in which subjects discriminate stimuli presented at fixation: the first 200 ms of brain activity is mostly bottom-up, not modulated by task constraints, followed by a second period of brain activity which is modulated by top-down, task related influences.

In previous studies, similarly to Philiastides et al. (2006) we described phase noise sensitivity of face ERPs in the time-window 100–300 ms (Rousselet et al., 2008b, 2009, 2010). Although in their second experiment (Philiastides et al., 2006) used only two noise levels, 30 and 45% phase coherence, their results suggest that when we use a larger range of noise levels, as in our previous experiments, noise sensitivity should be strongly reduced after 200 ms when it is made task irrelevant. We tested this hypothesis by asking subjects to perform two tasks: the same face identity discrimination task (face 1 vs. face 2) we used in previous studies (Rousselet et al., 2008b, 2009, 2010) and the same color discrimination task (red vs. green) used by Philiastides et al. (2006). We performed both group level and single-subject analyses to reveal the detailed time-course of the task effects.

## MATERIALS AND METHODS

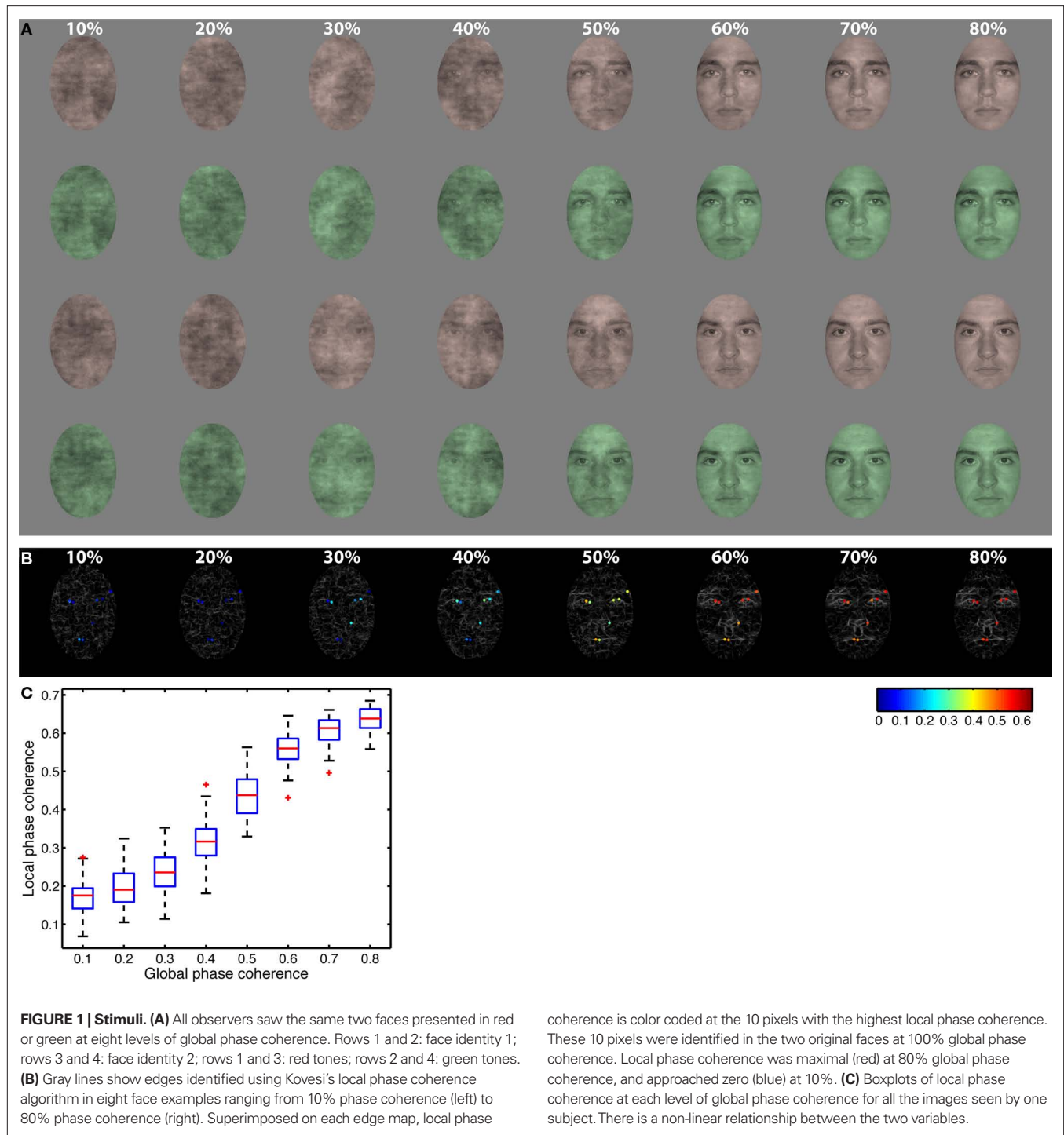
Square brackets indicate the boundaries of 95% confidence intervals (CIs) constructed using a percentile bootstrap with 1000 samples (Wilcox, 2005).

## SUBJECTS

We recruited 13 subjects, including the second and third authors, and 11 subjects from the Glasgow Psychology subject pool. Subjects' mean age was 24 years old (min = 20, max = 32); eight were females, 11 were right handed. Their mean high-contrast 63 cm decimal acuity was 104 (min = 99, max = 110); their low-contrast 63 cm decimal acuity was 96 (min = 89, max = 103). All subjects had a Pelli-Robson contrast sensitivity of 1.95 and successfully passed the Ishihara color blindness test for red-green color deficiencies. On average subjects had 19 years of education (min = 15, max = 23). All subjects except the two authors received £6/hour for their participation and all subjects gave written informed consent. The research ethics board from the University of Glasgow approved the research protocol.

## STIMULI

We used two front-view male face photographs cropped within a common  $4.3 \times 6.3$  oval frame and pasted on a uniform  $9 \times 9$  gray background (Figure 1). These faces were selected from a set of 10 faces, which are described in detail in previous publications (Gold et al., 1999; Husk et al., 2007). All stimuli had the same global amplitude spectrum. We added noise to their phase spectra so that their percentage of global phase coherence ranged from 10 to 80%, with 10% increments. Noise was random on each trial, which means that subjects never saw the exact same image twice. We also colorized the faces with red and green tones by manipulating



the hue ( $H$ ), saturation ( $S$ ), and value ( $V$ ) of the original images (red:  $H = 0.04$ ,  $S = 0.17$ ,  $V = \text{unchanged}$ ; green:  $H = 0.34$ ,  $S = 0.23$ ,  $V = \text{unchanged}$ ). The value ( $V$ ) was normalized so that, on average, each face regardless of color or phase coherence had the same average luminance (about  $33 \text{ cd/m}^2$ ) and RMS contrast (0.1). We colorized only the face itself and not the background, and used relatively small images to ensure that subjects paid attention to the face in the two tasks. This manipulation increases the likeli-

hood that task effects are due to a change in the task relevance of one stimulus dimension while subjects attempt to discriminate the same stimulus. A Dell Precision 390 workstation with Nvidia Quadro FX 3450/4000 graphics card and MATLAB Psychophysics Toolbox controlled the stimulus display. Images were displayed on a SAMSUNG SyncMaster 1100MB CRT monitor with a resolution of  $800 \times 600$  pixels and a 85-Hz refresh rate. The screen was  $28^\circ \times 21^\circ$  of visual angle.

## EXPERIMENTAL DESIGN

Testing was conducted in a sound-attenuated booth in which the monitor was the only source of light. An 80-cm viewing distance was maintained with a chinrest. We tested subjects in two experimental sessions. The first day was a practice behavioral session; the second day consisted of both behavioral tasks and simultaneous EEG recordings. Each day, subjects performed two interleaved tasks. On half of the blocks they performed a one-interval, two-alternative forced choice task discriminating between two faces. On the other half of the blocks, they discriminated between two colors. Identity and color tasks were blocked so subjects could focus on one task for an entire block of trials, without having to prepare to switch task on each trial (Johnson and Olshausen, 2003), in an attempt to increase the likelihood of finding strong task effects. The same stimuli were presented in the two tasks. In both tasks, on each trial, one face appeared briefly (36 ms), and subjects had to indicate which of two possible faces or two possible colors was presented by pressing 1 or 2 on the numerical pad of the keyboard. The association between button and identity/color was assigned randomly for all subjects. Subjects were given unlimited time to respond, and were told to emphasize response accuracy, not speed. All subjects performed the task with the same single pair of male faces throughout the experiment. Subjects saw eight conditions along a noise-signal continuum, from 10 to 80% phase coherence, with increments of 10% (Figure 1).

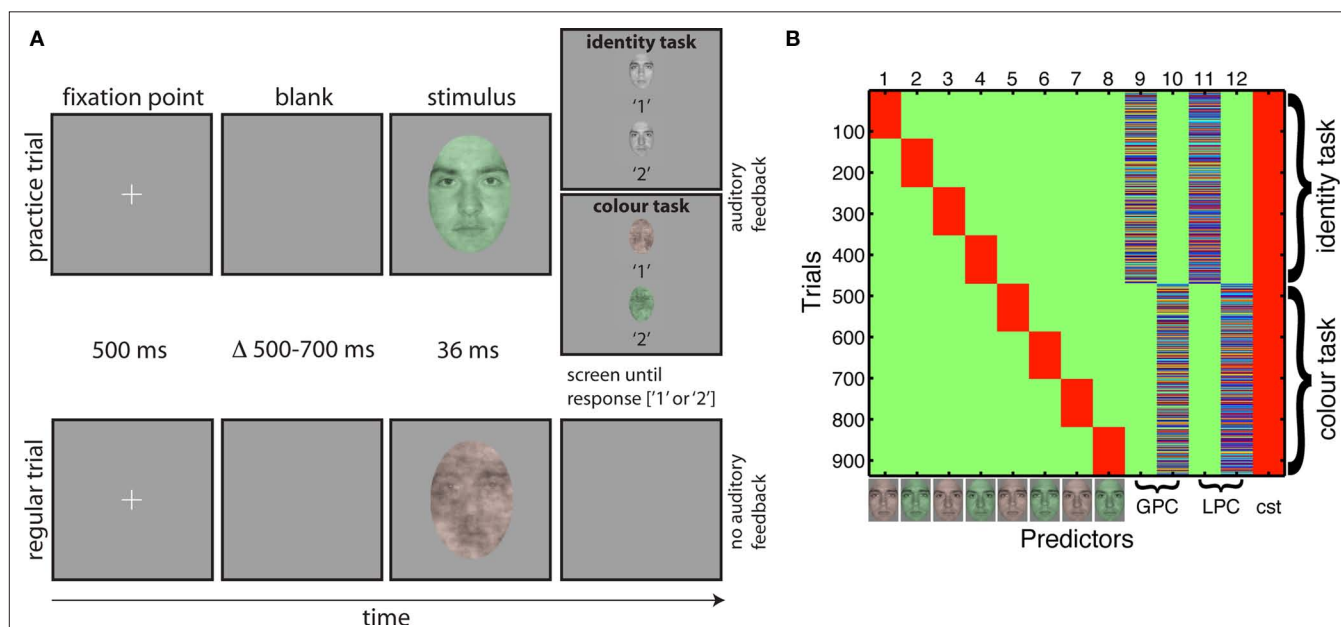
There were 10 blocks of 96 trials: 960 trials in total, with 120 trials per level of phase coherence. Within each block, there were equal repetitions of each face, each color and each phase coherence level.

Each block was preceded by 10 practice trials that allowed subjects to learn the stimulus-key association. Practice trials were used to ensure a high level of performance in older subjects, whose data are not reported here. In a regular trial, a small fixation cross – a 0.3 “+” in the middle of the screen – appeared for 500 ms, after which a blank screen was presented for a random duration ranging from 500 to 700 ms (Figure 2A). Then a test stimulus was presented for 36 ms, followed by a blank screen that stayed on until subjects provided their response. Practice trials were very similar, except that immediately after stimulus presentation, a choice screen appeared that showed each face in grayscale (identity task) or red and green noise textures (color task) simultaneously, one above each other with a corresponding label below each item. Auditory feedback was provided after the subject pressed a response key, with low- and high-pitched tones indicating incorrect and correct responses. Feedback was provided only during practice trials.

## EEG RECORDING AND PREPROCESSING

We acquired EEG data with a 128-channel Biosemi Active Two EEG system (BioSemi, Amsterdam, Netherlands). We recorded from four additional electrodes – UltraFlat Active BioSemi electrodes – below and at the outer canthi of both eyes. Analog signal was digitized at 512 Hz and band-pass filtered online between 0.1 and 200 Hz. Electrode offsets were kept between  $\pm 20 \mu\text{V}$ .

Offline, data were average-referenced. Then, we removed bad channels without interpolation, applied a 40-Hz low-pass filter and epoched the data between –300 and 1,200 ms. An ICA (Makeig et al.,



**FIGURE 2 | Tasks and design matrix. (A)** Organization of practice trials and regular trials in the two tasks. A trial started with the presentation of a fixation point for 500 ms. Then, after a random delay ranging from 500 to 700 ms, a stimulus was presented for about 36 ms. During practice trials (top row), a choice screen appeared immediately after the stimulus, showing the two targets of the task and their associated response keys. The screen stayed on until the subject's response, which was followed by auditory feedback, before the trial sequence resumed. During regular trials (bottom row), a blank screen appeared

immediately after the stimulus, and remained on until the subject's response. No feedback was provided during regular trials. Stimuli are not drawn to scale. **(B)** Example of a design matrix in one subject (color scale: green = 0; red = 1). The first eight predictors were categorical: they indicate the stimulus type (i.e., red or green, face 1 or 2) and the task. The next four predictors were continuous: global phase coherence (GPC) and local phase coherence (LPC) in the identity and color tasks. Each continuous predictor was z-scored independently before insertion in the design matrix. The last column was a constant term (cst = 1s).



2004), as implemented in the *runica* EEGLAB function (Delorme and Makeig, 2004; Delorme et al., 2007) was then computed and we removed components corresponding to blink activity, identified by visual inspection of their scalp topographies, time-courses and activity spectra. Subsequently, we re-epoched the data between  $-300$  and  $500$  ms, and subtracted the average baseline activity from each time point. Trials with abnormal activities were excluded based on a  $\pm 100$ - $\mu$ V threshold for extreme values. An epoch was rejected for abnormal trend if it had a slope larger than  $75$   $\mu$ V/epoch and a regression  $R^2$  larger than  $0.3$ . All remaining trials were included in the analyses, whether they were associated with correct or incorrect behavioral responses. After epoch rejection, the average number of trials per subject was  $904$  (min =  $849$ , max =  $958$ ).

### GENERAL LINEAR MODELING OF EEG DATA

Subjects' epoched data were modeled using LIMO EEG, an open source Matlab toolbox for hierarchical GLM, compatible with EEGLAB: [https://gforge.dcn.ed.ac.uk/gf/project/limo\\_eeg/](https://gforge.dcn.ed.ac.uk/gf/project/limo_eeg/) (Pernet et al., 2011). The general linear model was used to express single-trial ERP amplitudes, in microvolt, independently at each time point and each electrode, using the model:

$$\begin{aligned} \text{ERP}_{t,e} &= \mathbf{X}\mathbf{B} + \epsilon \\ \text{ERP}_{t,e} &= \beta_0 + \text{cat}_1\beta_1 + \text{cat}_2\beta_2 + \dots + \text{cat}_8\beta_8 + \phi_{\text{G-ID}}\beta_9 + \phi_{\text{G-CO}}\beta_{10} \\ &\quad + \phi_{\text{L-ID}}\beta_{11} + \phi_{\text{L-CO}}\beta_{12} + \epsilon \end{aligned} \quad (1)$$

In Eq. 1, all trials for each time frame  $t$  and electrode  $e$  ( $\text{ERP}_{t,e}$ , dimension  $1 \times n$ ) were modeled as the sum of a constant term  $\beta_0$ , the eight experimental conditions (each combination of stimulus identity, color and task  $-2 \times 2 \times 2 = 8$ ,  $\text{Cat}_{1-8}$ —corresponding to the first 8 columns of the design matrix), the global phase coherence in the identity and the color tasks ( $\phi_{\text{G-ID}}, \phi_{\text{G-CO}}$ ), the local phase coherence in the identity and the color tasks ( $\phi_{\text{L-ID}}, \phi_{\text{L-CO}}$ ), and an error term  $\epsilon$ . All predictors formed the design matrix  $\mathbf{X}$  of dimension  $n \times p$  (Figure 2B,  $p = 13$ ). The beta parameters (dimension  $p \times 1$ ) were found using an ordinary least square solution.

Global phase coherence was our image noise manipulation. Kovess's (1999, 2003) local phase coherence is a measure of wavelet phase alignment across spatial frequencies, which is independent of image contrast and luminance. Local phase coherence may predict subjects' behavior in a natural scene classification task (Gaspar and Rousselet, 2009) and seems to provide a good representation of non-linear changes in local image structure imposed by the linear global phase coherence manipulation (Rousselet et al., 2008b, 2009, 2010). In our stimuli, pixels with high local phase coherence corresponded to local edges around the eyes, nose, and mouth (Figure 1).

The design matrix represents a typical ANCOVA model with categorical and continuous predictors. However, whereas in ANCOVA one is usually interested in the categorical effects whilst controlling for covariates, here we were interested in the covariate effects: we looked at the relationship between image phase coherence and single-trial ERP amplitude in the identity and the color tasks whilst accounting for the main effects of identity, color, and task. For instance, the average ERP in one condition (e.g., identity discrimination of green face 1) could differ from the average ERP in another condition (e.g., color discrimination of green face 1)

because subjects might be less attentive in the easier color task compared to the more challenging identity task. These confounding mean ERP differences were accounted for in the design matrix, thus allowing us to measure how single-trial ERPs were modulated by image noise in the two tasks. We used linear contrasts to combine the beta weights associated with the global and local phase coherence predictors in the identity task (column 9 + column 11 in Figure 2B) and in the color task (column 10 + column 12) to study the time-course of the overall ERP noise sensitivity:

$$d_{\text{identity}} = \beta_9 + \beta_{11} \quad (2)$$

$$d_{\text{color}} = \beta_{10} + \beta_{12} \quad (3)$$

We did not look at task modulations separately for global phase coherence and local phase coherence because these two predictors were strongly correlated (identity: mean angle =  $17.4$  [ $17.3, 17.6$ ], mean correlation =  $95.4$  [ $95.3, 95.5$ ]; color: mean angle =  $17.6$  [ $17.3, 17.9$ ], mean correlation =  $95.3$  [ $95.1, 95.5$ ]; mean angle differences between the two tasks =  $-0.1727$  [ $-0.5705, 0.2544$ ]). High correlation between regressors may lead to unstable beta parameter estimates, whereas their linear combination remains stable, hence our analysis of the combination of global and local phase coherences. We refer to this summary statistics as noise sensitivity in the rest of the paper and explored task effects by contrasting  $d_{\text{identity}}$  and  $d_{\text{color}}$ .

### GROUP LEVEL ANALYSES

Group analyses of noise sensitivity task modulations were computed using a bootstrap- $t$  technique for paired samples with 1000 resamples (Wilcox, 2005). Although full scalp analyses are possible in LIMO EEG, we performed the analyses at only one electrode for two reasons: first because we observed in previous studies that noise sensitivity is localized at few posterior electrodes that display redundant information (Rousselet et al., 2008b, 2009, 2010); second because we wanted to compare different group analyses to single-subject analyses. We analyzed group results using four different ways to pull data together.

A popular way to do group analyses is to average the data across subjects, find the best electrode in this group average, and make a measurement at that same electrode in all subjects—group defined best electrode. Here, we averaged across subjects the  $R^2$  maps of the ANCOVA model fit to the data, and selected the electrode showing the largest mean  $R^2$ . Bootstrap paired  $t$ -tests were then computed between noise sensitivity contrasts in the identity and color tasks at all time points at this electrode.

A potentially more fruitful way to do group statistics is to optimize the electrode by selecting the best electrode independently in each subject (Foxe and Simpson, 2002; Liu et al., 2002; Rousselet et al., 2010). We thus took the electrode at which the model provided the best fit independently in each subject, i.e., where  $R^2$  was the largest for each subject— $R^2$  optimized electrode. Then we computed paired  $t$ -tests between noise sensitivity contrasts from these potentially spatially different electrodes. The signal at  $R^2$  optimized electrodes was the most sensitive to image and task parameters as described by the design matrix and therefore constitutes the most likely candidate for reflecting the activity of cortical sources sensitive to image information. Hence, this kind of optimized

averaging tends to average signals that reflect common processing across subjects, whereas using the same spatial electrode may lead to averaging signals reflecting different processes.

Yet another way to optimize electrodes across subjects consists in selecting for each subject the electrode with the largest noise sensitivity task difference –task effect optimized electrode. In this case, instead of taking the electrode where the ANCOVA model provided the best fit, we selected for each subject the electrode showing the strongest noise sensitivity task effect. The paired *t*-test was then computed between noise sensitivity contrasts from these potentially spatially different electrodes.

Finally, we used the maximum absolute beta coefficients across electrodes computed at each time point (the envelope), to ensure our analyses did not miss local maxima at electrodes other than the one showing the largest  $R^2$ . For every subject, we computed a paired *t*-test between noise sensitivity contrasts from the envelopes.

For both group and single-subject analyses, task modulations at one electrode were quantified by normalizing the maximum absolute task difference in noise sensitivity by the maximum absolute noise sensitivity in the identity task –the maxima were defined across time frames:

$$\text{Task modulation} = 100 \times \max(|d_{\text{identity}} - d_{\text{color}}|) / \max(|d_{\text{identity}}|) \quad (4)$$

### CONTROL FOR MULTIPLE COMPARISONS

We controlled for multiple comparisons using bootstrap and the clustering technique as implemented in the Matlab Fieldtrip toolbox, with a minimum of two neighboring channels per cluster (Maris and Oostenveld, 2007). As described in (Pernet et al., 2011), the clustering technique in LIMO EEG works for analyses both at single electrodes (temporal clustering) and at multiple electrodes (spatial–temporal clustering). For group analyses, because only one electrode or equivalent electrode was considered, we employed temporal clustering to control for multiple comparisons. For single-subject analyses, because the whole scalp was analyzed, we employed spatial–temporal clustering (familywise error rate = 0.05).

For *t*-tests and ANOVAs the validated bootstrap technique includes centering the empirical distributions of each between-subject and within-subject levels so that the null-hypothesis of no difference in means is true (Berkovits et al., 2000; Wilcox, 2005; Seco et al., 2006). Thus, for the group paired *t*-tests, noise sensitivity contrasts across subjects were centered for each condition separately and paired *t*-tests were computed 1000 times by sampling subjects with replacement (Wilcox, 2005). However, this technique is not appropriate to our ANCOVA single-subject analyses because the continuous covariates can potentially have as many levels as trials. We used therefore a different strategy to derive an estimate of the sampling distribution of our *F* statistics under the null-hypothesis. For each subject, epoched single-trials were sampled with replacement and fitted to the original design matrix, thus breaking the link between the data and the design –an estimation of the data distribution under the null-hypothesis  $H_0$ .

For both the bootstrap paired *t*-test (group analyses) and the bootstrap ANCOVA (single-subject analyses), in each bootstrap loop we first computed the sum of each temporal or spatial–temporal cluster

of contiguous significant *F* values (univariate  $p < 0.05$ ), separately for each predictor, each linear contrast, and in the case of the ANCOVA, for the global fit of the entire model ( $R^2$ ). Second, we saved the maxima across these cluster sums – one maximum for each *F* test (familywise correction). After performing these steps 1000 times for group statistics and 600 times for single-subject analyses (as recommended for various linear models, Wilcox, 2005), we used the 95th percentiles of the bootstrapped maximum *F* cluster sums to threshold the original *F* cluster sums. For each test, the significant original *F* values (univariate  $p < 0.05$ ) were clustered and if their sum were larger than the corresponding bootstrapped maximum cluster sum threshold, the cluster was significant.

### NOISE SENSITIVITY CLUSTER STATISTICS

For each subject, we used a percentile bootstrap rather than using an *F* test of noise sensitivity (sum of beta coefficients for global and local phase coherence). Bootstrap distributions were used to compute 95% CIs under  $H_0$ , during the same simulation that was used to estimate the *F* distributions of the ANCOVA parameters. These thresholds were then applied to each bootstrap to mark significant noise sensitivity. Significant effects were then clustered and a maximum sum of absolute noise sensitivity was saved for each bootstrap. The bootstrap distributions of maximum sum of absolute noise sensitivity computed under  $H_0$  were used to cluster the observed noise sensitivity in each task.

### SHIFT FUNCTION ANALYSES OF THE DECILES OF SINGLE-TRIAL ERP DISTRIBUTIONS

We also used the shift function to measure how single-trial ERP distributions changed from the identity task to the color task. The shift function compares entire distributions instead of relying exclusively on one point estimate such as the mean or the median. In our application of the shift function, the *x*-axis is the Harrell–Davis (hd) estimator of quantiles one to nine of the single-trial ERPs in the identity task (see Wilcox, 2005, pp. 71–73 and 139–141). The *y*-axis is the difference, Delta, between the hd estimators of the quantiles of the identity and color ERP distributions. Hence, the shift function represents how much the data from one task must be shifted to be comparable to the data from another task at each quantile. Task differences were estimated by a bootstrap procedure, and corrected for multiple comparisons such that the simultaneous probability coverage of the 9 CIs remained close to the nominal 0.05 alpha level (see Wilcox, 2005, pp. 151–155). The analyses were performed on modeled single-trial ERP data at the max  $R^2$  electrode (i.e., the electrode at which the model explained best the data); they included all the significant time points that contained the maximum noise sensitivity task difference. Modeled ERP are more meaningful to analyze because they are reconstructed after removing the error term, the part of variance that the model cannot explain.

### RESULTS

We consider first the group analyses, second the single-trial analyses, third the comparison of group and single-trial analyses and fourth the shift function analyses of the deciles of the single-trial ERP distributions.

## GROUP ANALYSES

### Group defined best electrode

If the best electrode is defined as the electrode showing the largest mean  $R^2$  across subjects, we obtain the results in **Figure 3A**. This best electrode was right posterior–lateral (B8 in the Biosemi system, between PO8 and PO10) and had a maximum mean  $R^2$  of 0.23 [0.17, 0.32] that peaked at 141 ms. The mean ERPs at the eight global phase noise levels started to diverge shortly after 100 ms in the identity and the color tasks. The parametric ERP modulation by image noise can be better appreciated by looking at the time-course of the group-averaged noise sensitivity, which peaked at the same electrode and time point as  $R^2$  did. Noise sensitivity was reduced in the color task compared to the identity task in a single cluster, between 139 and 277 ms after stimulus onset (**Table 1**). At the latency of the maximum task effect, 242 ms, there was 20.7% noise sensitivity reduction compared to the maximum sensitivity in the identity task.

### $R^2$ optimized electrode

The electrode with the largest  $R^2$  was also the electrode with maximum noise sensitivity or was part of the same cluster as the electrode with maximum noise sensitivity and behaved similarly to it. Across subjects, max  $R^2$  electrodes were all located in a cluster of lateral posterior electrodes, as reported in previous experiments and as expected from the face literature.  $R^2$  averaged across subjects was stronger over the right hemisphere. This pattern was also found in individual subjects (**Figures 5–8**): eight subjects had a maximum  $R^2$  at right hemisphere electrodes; two subjects at left hemisphere electrodes; three subjects at midline electrodes. The right hemisphere electrodes of maximum model fit included B8 (one subject) or one of its neighbors (seven subjects). The maximum mean  $R^2$  at the optimized electrode was 0.27 [0.2, 0.35], and peaked at 139 ms (**Figure 3B**; **Table 1**). As expected if  $R^2$  results were sufficiently consistent across subjects, this optimized maximum average  $R^2$  was larger than at the group defined best electrode. There was a significant task effect in a single cluster between 172 to 275 ms post-stimulus onset, with 18.7% noise sensitivity reduction at the latency of the maximum task difference, 213 ms.

### Task effect optimized electrode

Results of this analysis are presented in **Figure 3C**. Although we selected for each subject the electrode showing the largest task difference in noise sensitivity, no differences could be observed at the group level. Indeed, taking the largest effect can be misleading because certain predictors can be significant at electrodes and time frames at which the overall model does not explain the data significantly. This is indeed what we found: except for 3 subjects for whom maximum task effects occurred at electrodes that were part of the cluster of electrodes with the maximum  $R^2$ , for the other 10 subjects  $R^2$  was lower and early noise sensitivity (<200 ms) was weak at the electrodes of maximum task differences (**Figure 3C**). Hence, across subjects, noise sensitivity and task effects were not significant at the electrode optimized based on maximum task differences.

### Maximum absolute betas

Analyses on the beta coefficient *envelopes* gave results similar to those obtained on the group defined best electrode and the  $R^2$  optimized electrode (**Figure 3D**; **Table 1**). Two significant clusters

were observed: a first task effect occurred between 154 to 254 ms post-stimulus onset, with 18.2% noise sensitivity reduction at the latency of the maximum task difference, 197 ms; a second effect of similar size occurred after 400 ms. This analysis suggests that we did not miss the big picture by defining the electrode to analyze based on the group-averaged  $R^2$  or the single-subject max  $R^2$ . The statistical tests might suggest that the group defined best electrode did a better job because it showed a significant task effect earlier than the one observed in the  $R^2$  optimized test. However, picking the best group electrode to show group effects is circular because what the result ought to be is unknown. By contrast, selecting the best electrode separately in each subject takes into account inter-subject variability and leads to group results more sensitive to individual differences.

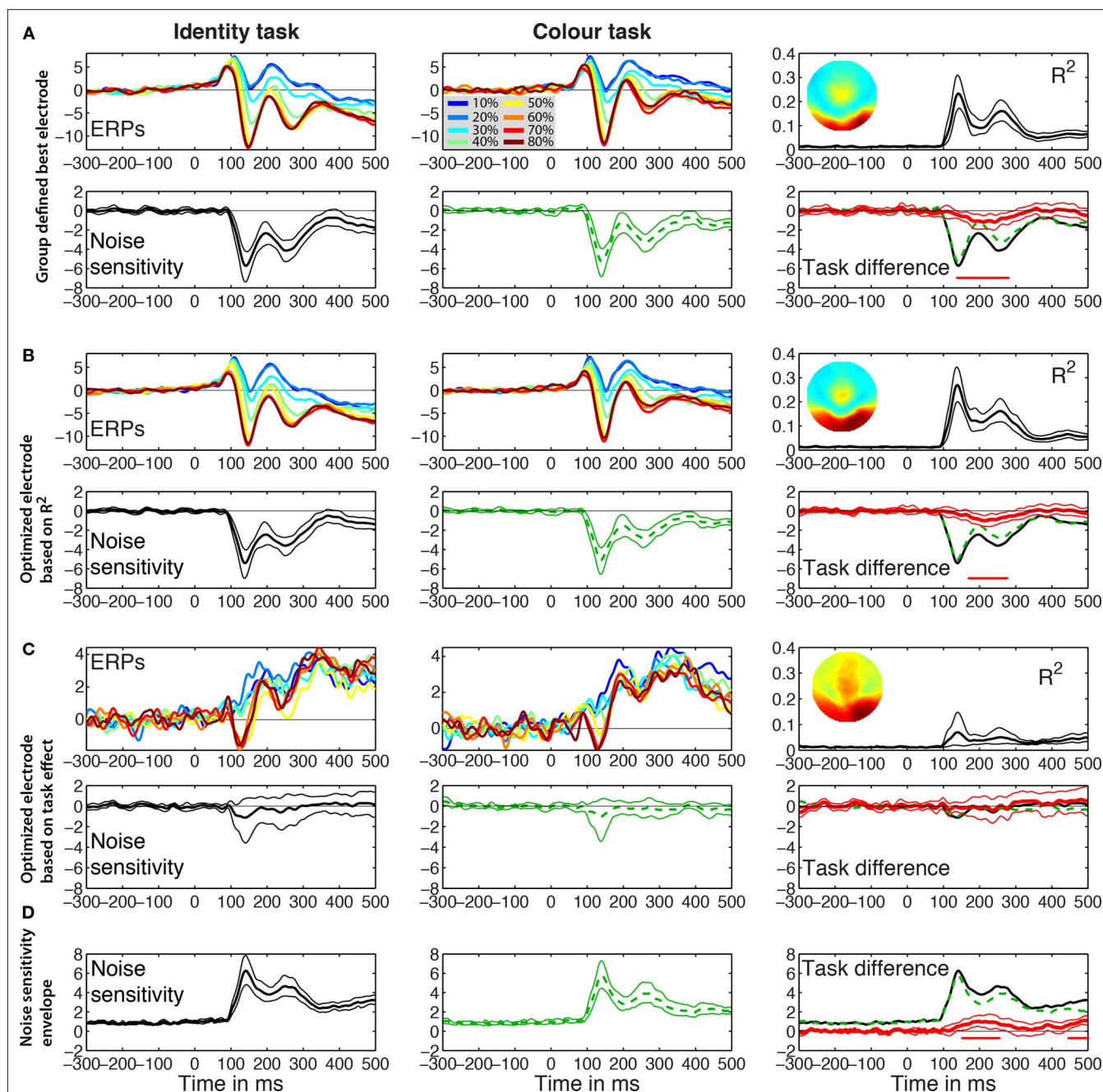
In addition to group analyses at all time points, we compared noise sensitivity across subjects at the latency of the P1, N170, and P2 peaks, for the sake of comparisons with previous studies (**Figure 4**). For each subject, the latency of the peaks was measured at the max  $R^2$  electrode in the 80% coherence condition (face ERP). Measurements at electrode B8 gave similar results. Then, the mean sensitivity was measured in time-windows encompassing five time points on either side of the peak latency, hence about 21.5 ms. This analysis revealed, across subjects, weak to no noise sensitivity around the P1 (~96 ms), and stronger noise sensitivity around the N170 (~146 ms) and the P2 (~207 ms). Importantly, only the P2 showed task modulations of noise sensitivity. Similar P2 results were obtained at the latency of the P2 defined at the group level.

## SINGLE-SUBJECT ANALYSES

**Figures 5–8** provide, for each subject, a detailed description of their  $R^2$ , noise sensitivity, task effects and behavioral results. The time-course of the  $R^2$  functions and of the beta coefficients for noise sensitivity are similar to those reported previously in young subjects (Rousselet et al., 2008b, 2009, 2010). Because the main purpose of our study was to quantify task modulations of early ERP noise sensitivity, we focus the report of the single-trial analyses on the electrode showing the maximum  $R^2$  for each subject. All these electrodes were found at posterior–lateral locations. No comparable fits were observed at frontal electrodes. Thus, our analyses seem to capture task modulations of evoked noise sensitivity from the visual system, rather than electrophysiological correlates of the top-down modulation signal itself. Figures at the electrodes showing the maximum noise sensitivity in the identity task or the color task were almost identical to those presented here (max  $R^2$ ) because these electrodes were either the same electrodes or part of the same cluster of electrodes.

Single-trial analyses revealed an inter-subject variability hidden behind the seemingly simple group averages and statistics. Individual subjects differed widely in the shape of their ERPs,  $R^2$  functions, scalp topographies, nature and time-course of the task effects (**Figures 5–8**). The mean of the maximum  $R^2$  measured in each subject was 0.31 [0.24, 0.38], min = 0.17, max = 0.57; it peaked at 141 ms [136, 149], min = 133, max = 186. Image noise sensitivity started at about 100 ms in both tasks. In the identity task, the median onset was 100 ms [98, 105], min = 92, max = 115; in the color task it was 100 ms [96, 111], min = 86, max = 133. The median difference between the two tasks was 0 ms [−7.8, 4],





**FIGURE 3 | Event-related potentials group results.** The left and middle columns contain the results for the identity and the color tasks. The right column shows the  $R^2$  and the noise sensitivity task differences. **(A)** Group defined best electrode. Mean ERPs are shown color coded at each level of global phase coherence in the two tasks. In the  $R^2$  plot, the inset shows the topographic map of the interpolated  $R^2$  values at the latency of the maximum  $R^2$ . Noise sensitivity is the sum of the global and local phase coherence beta weights in  $\mu\text{V}/\text{std}$  of the predictor. Thick lines represent averaged data, surrounded by thin lines for the 95% percentile bootstrap CI. The red horizontal bar under the zero line indicates time points of significant effects, based on bootstrap  $t$ -test with temporal cluster correction for multiple comparisons. The task difference in red is identity (black continuous line) minus color (green dashed line). At the latency of the maximum task

difference observed within the first 300 ms after stimulus onset, noise sensitivity in the identity task was  $-4 \mu\text{V}/\text{std}$ , color =  $-2.8 \mu\text{V}/\text{std}$ , difference =  $-1.2 \mu\text{V}/\text{std}$ . **(B)**  $R^2$  optimized electrode. The topographic map was obtained by averaging the maps from individual subjects. At the latency of the maximum task difference, noise sensitivity in the identity task was  $-2.8 \mu\text{V}/\text{std}$ , color =  $-1.8 \mu\text{V}/\text{std}$ , difference =  $-1 \mu\text{V}/\text{std}$ . **(C)** Task effect optimized electrode. The  $R^2$  bump between 100 and 200 ms was mostly due to three subjects (S9–S11) who had maximum task effects at electrodes that were part of the cluster of electrodes with the maximum  $R^2$ . **(D)** Noise sensitivity envelope. The maximum across electrodes of the absolute noise sensitivity was used for each subject. At the latency of the maximum task difference, noise sensitivity in the identity task was  $3.9 \mu\text{V}/\text{std}$ , color =  $2.8 \mu\text{V}/\text{std}$ , difference =  $1 \mu\text{V}/\text{std}$ .

min = −21, max = 16. Results using the mean were similar and a shift function analyses failed to reveal significant differences at any deciles of the distribution of onset differences.

Only eight subjects showed a significant task modulation of noise sensitivity. Noise sensitivity decreased in the color task compared to the identity task in seven subjects:  $S_1 = 29\%$ ,  $S_3 = 46\%$ ,  $S_4 = 35\%$ ,  $S_6 = 19\%$ ,  $S_7 = 39\%$ ,  $S_{10} = 29\%$ ,  $S_{11} = 39\%$ , mean = 33.7%. Subject  $S_2$  showed an effect in the opposite direction (**Figure 5**), with significantly stronger noise sensitivity in the color task than in the identity task (40.7% sensitivity increase). All subjects but  $S_{11}$  (**Figure 7**) had a single cluster of significant task differences.

There was no significant link between task effects and behavioral thresholds: mean behavioral 75% correct threshold was 0.37 global phase coherence for the five subjects who did not show a significant ERP task modulation, it was 0.39 global phase coherence for subjects

who did show a reduction in the color task (difference = −0.02 [−0.07; 0.04]), and 0.37 global phase coherence for  $S_2$ , who showed opposite ERP results. Analyses with the median gave similar results.

The max  $R^2$  electrodes almost never showed significant differences between the two faces (identity task) or the two colors (color task), in keeping with previous reports on faces, cars, and words (Nobre et al., 1998; Philiastides et al., 2006; Rousselet et al., 2008b, 2010). One subject showed both significant identity sensitivity in the identity task and a significant task modulation of identity sensitivity at few time points around 450 ms after stimulus onset. Five subjects showed either identity or color sensitivity in one or the other task, but without significant task modulation, or significant task modulations but without significant identity or color sensitivity.

GROUP VS. SINGLE-TRIAL ANALYSES

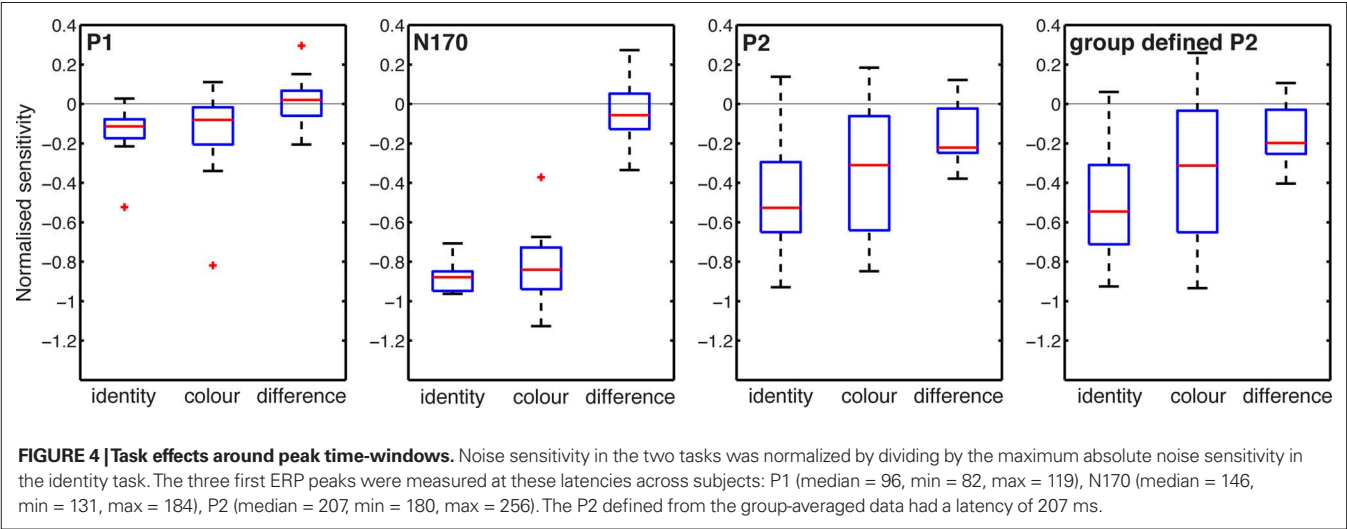
Group analyses suggest a decrease in noise sensitivity in the color task compared to the identity task around 140–300 ms post-stimulus onset. This task modulation was observed at the group defined best electrode and the  $R^2$  optimized electrode (**Figure 9**). In the spatial-temporal clusters containing these electrodes, single-trial analyses revealed a different picture: 8 subjects out of 13 had a significant task modulation of noise sensitivity; 1 had an increased sensitivity in the color task and 7 had a decreased sensitivity. In the time-window of the group effect, only six subjects showed a significant effect; a maximum of five subjects showed an effect simultaneously, including the subject who had an effect in the opposite direction. Thus, in our sample, at any time point showing a significant group effect, there were at most 5 subjects out of 13 showing a significant effect, 4 of which in the same direction as the group effect (31%).

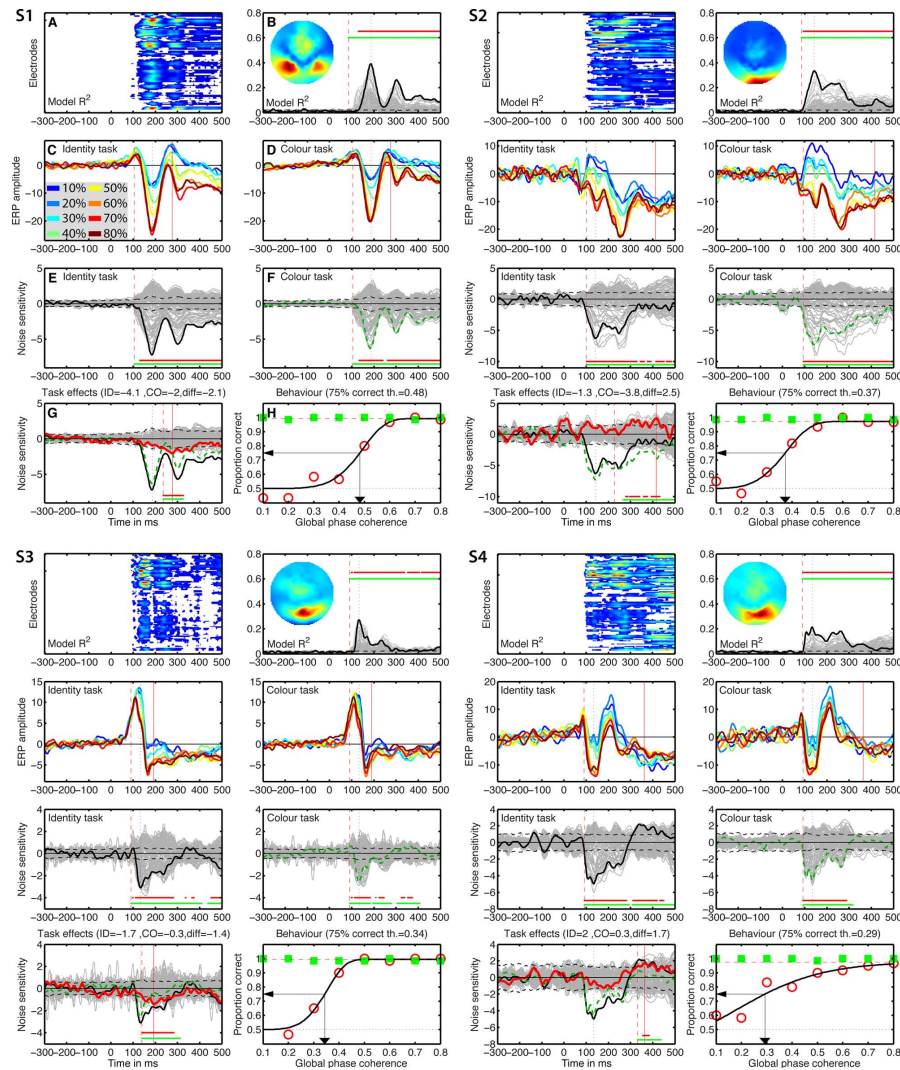
Onsets and durations of task effects also revealed discrepancies between group and single-trial analyses (**Table 1**). At the  $R^2$  optimized electrode, the group task effects started at 172 ms and lasted 103 ms. In contrast, for the eight subjects that showed significant effects, the average task effect onset was 214 ms [155, 271] min = 86, max = 332; the average task effect duration was the 186 ms [142, 234], min = 92, max = 320.

Table 1 | Task effects.

	Group analyses			Single-trial analyses
	Best group electrode (B8)	$R^2$ optimized electrode	Beta envelope	Mean of eight subjects with significant task effects
Onset	139 ms	172 ms	154 ms	214 ms [155, 271]
Peak latency	242 ms	213 ms	197 ms	297 ms [246, 352]
Effect size	20.7%	18.7%	18.2%	34.6% [28.9, 39.5]
Duration	139 ms	104 ms	101 ms	186 ms [142, 234]

Peak latency is the latency of the maximum absolute task effect. Effect size is defined in Section “Group Level Analyses,” Eq. 4.

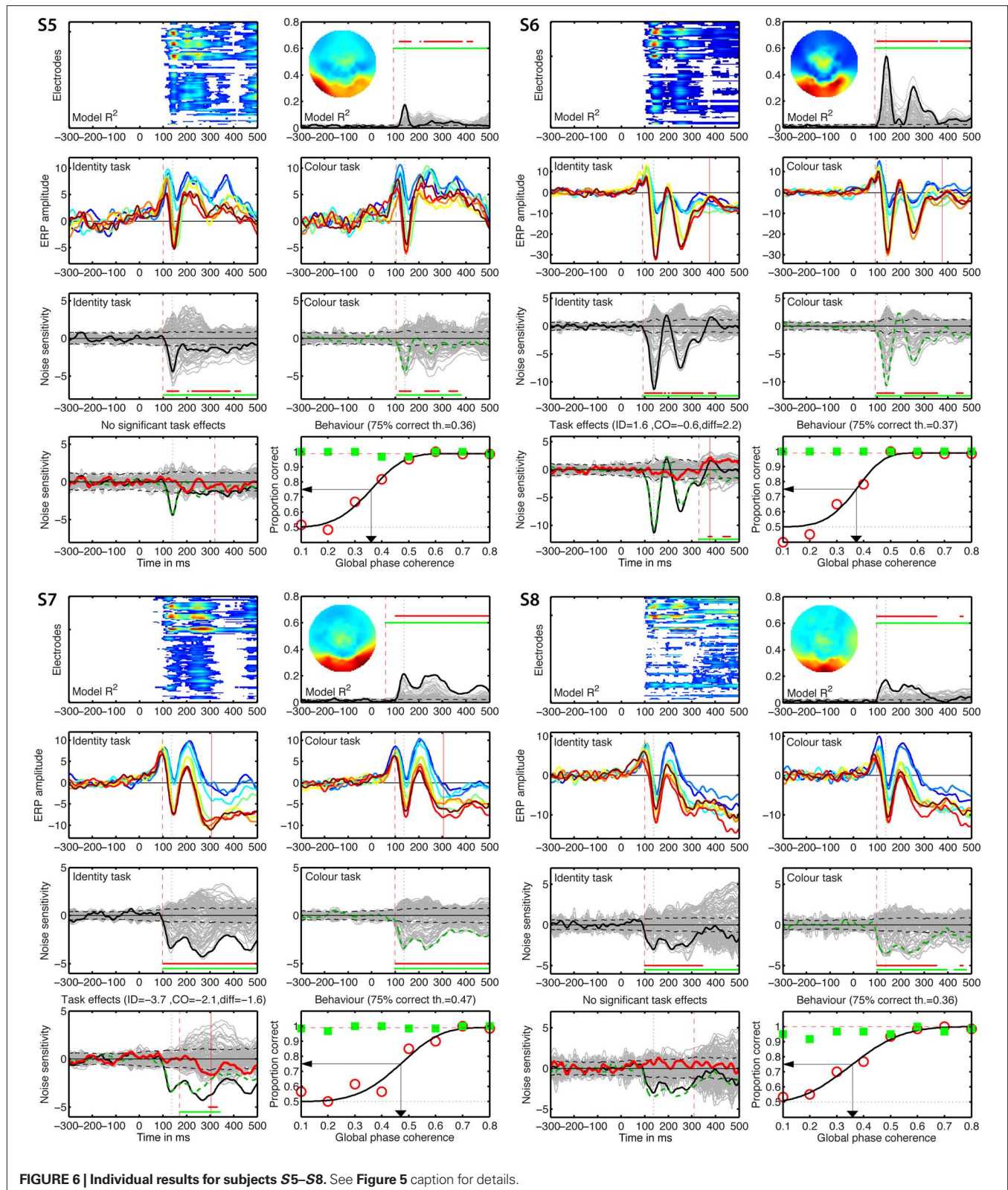




**FIGURE 5 | Individual results for subjects S1 to S4. (A)** Statistically significant model  $R^2$  at all the electrodes and time frames from  $-300$  to  $500$  ms after stimulus onset. Electrodes are stacked up along the y-axis. The tick on the y-axis marks the electrode at which the maximum  $R^2$  was recorded. This electrode is plotted as a continuous black line in **(B)**.  $R^2$  ranges from near zero in blue to the maximum for that subject in red. **(B)** Model  $R^2$  at all the electrodes and time frames. The electrode at which the maximum  $R^2$  was recorded is plotted in black. The other electrodes are plotted in gray. The inset shows the topographic distribution of the  $R^2$  at the latency of the maximum, indicated by a vertical black dotted line. This vertical line is also plotted in all the other panels for comparisons. For subject  $S_1$ ,  $R^2$  had a bilateral occipital-lateral distribution, with a maximum over the left hemisphere (left bottom red cluster). The red vertical dashed line indicates the time frame of the earliest significant  $R^2$  across all electrodes. Near the top of the panel, the upper horizontal line (red) marks significant time frames at the maximum  $R^2$  electrode. The lower horizontal line (green) marks significant time frames of the spatial-temporal cluster to which the maximum  $R^2$  electrode belonged. For subject  $S_1$ , this horizontal line starts at the latency of the earliest significant model fit (red vertical dashed line), indicating that the maximum  $R^2$  electrode is part of a spatial-temporal cluster that captures the earliest effects. The horizontal dashed line is the univariate one-sided 95% CI of the  $R^2$  under  $H_0$ , at the maximum  $R^2$  electrode. Although this is for illustration only, because the actual statistical test was based on spatial-temporal clusters, it gives a good indication of the  $R^2$  values expected by chance. **(C)** Mean ERPs in the identity task. The red vertical dashed line indicates the time frame of the earliest significant noise sensitivity in the identity task, across all electrodes. This line is also plotted in **(E)**. The red continuous vertical line indicates the latency of the

maximum task difference and is also plotted in **(D,G)**. **(D)** Mean ERPs in the color task. The red vertical dashed line indicates the time frame of the earliest significant noise sensitivity in the color task, across all electrodes. This line is also plotted in **(F)**. **(E)** Noise sensitivity beta coefficients in the identity task. Noise sensitivity at the maximum  $R^2$  electrode is plotted in black, the other electrodes in gray. Units are  $\mu V$ /std of the predictor. Near the bottom of the panel, the upper horizontal line (red) marks significant noise sensitivity time frames at the maximum  $R^2$  electrode. The lower horizontal line (green) marks significant noise sensitivity time frames of the spatial-temporal cluster to which the maximum  $R^2$  electrode belonged. The black horizontal dashed lines show the univariate two-sided 95% confidence interval of noise sensitivity under  $H_0$ , at the maximum  $R^2$  electrode. **(F)** Noise sensitivity beta coefficients in the color task. Noise sensitivity at the maximum  $R^2$  electrode is plotted as a green dashed line, the other electrodes in gray. **(G)** Noise sensitivity beta coefficient task differences. Noise sensitivity differences at the maximum  $R^2$  electrode are plotted as a thick red line, the other electrodes in gray. The black continuous line and the green dashed line are the same as those in **(E,F)**. The red continuous vertical line indicates the latency of the maximum task difference. At that latency, the title indicates the amplitude of the noise sensitivity in the identity task (ID), in the color task (CO), and the difference between the two tasks (diff). **(H)** Proportion correct as a function of global phase coherence, in red circles for the identity task, in green squares for the color task. Data from the identity task were fitted with a cumulative Weibull function. The vertical arrow points to the 75% correct threshold in the identity task. The threshold appears in bracket in the title. The red horizontal dashed line marks the maximum proportion correct in the identity task obtained from the fit.

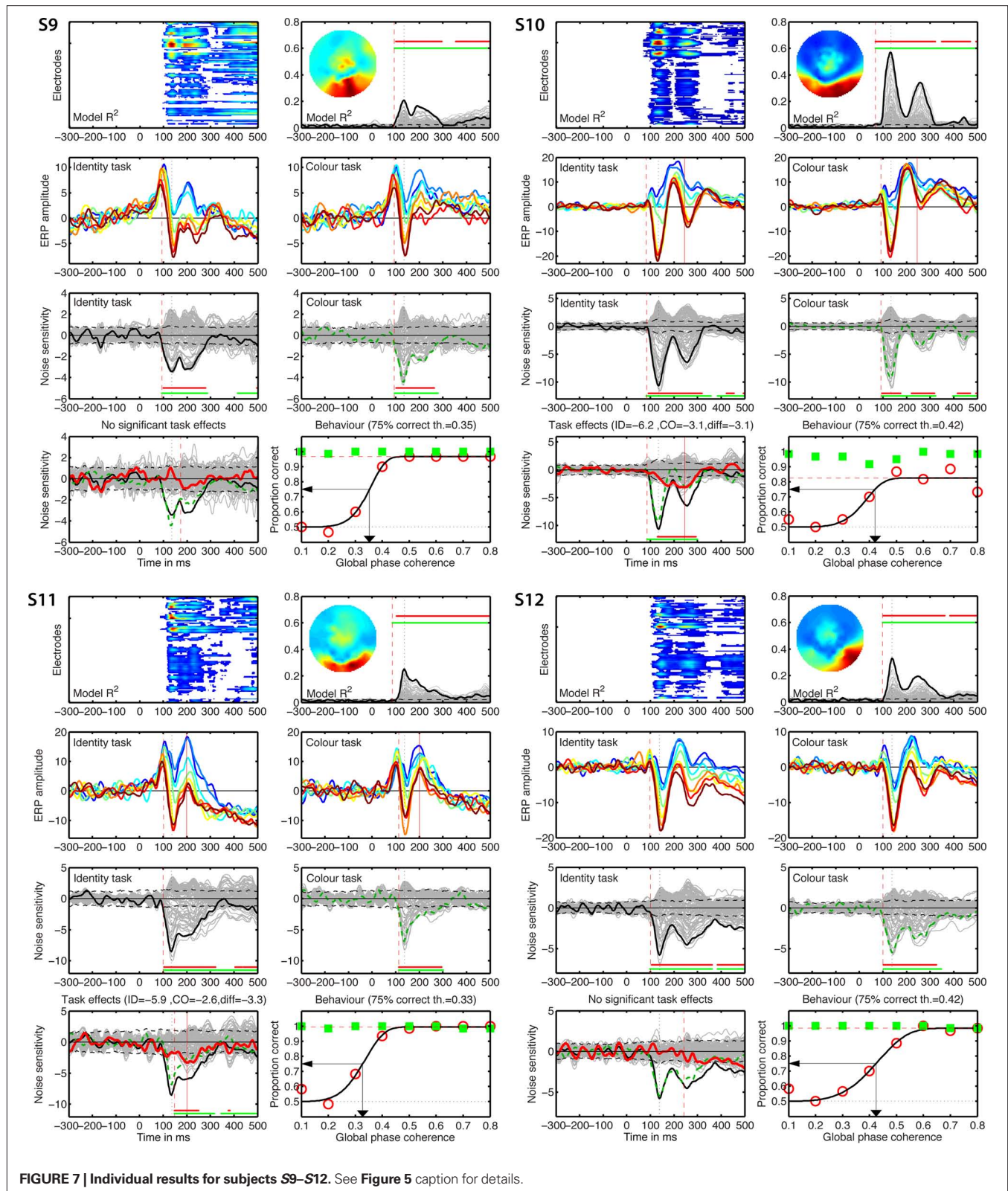




Single-trial results can be misleading too. The results at the group defined best electrode showed significant effects of longer duration than results at the  $R^2$  optimized electrode (Figure 9).

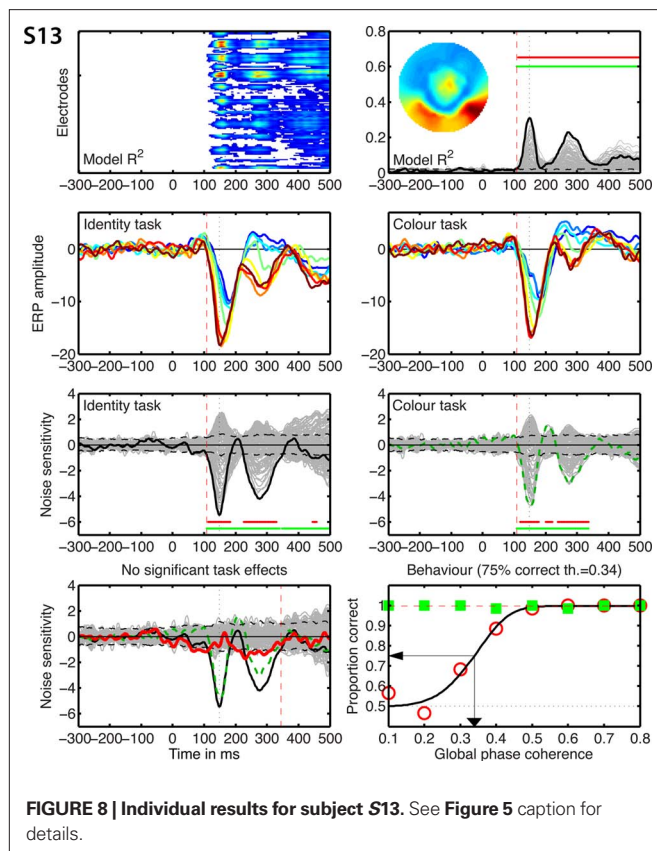
However, results at the group defined best electrode mix together electrodes that do not necessarily provide the best fit in all subjects. This means that some of the effects at this electrode are





not as meaningful as the results observed at the best fitting electrodes. In particular, the late task modulations observed in the left column of Figure 9, correspond to clusters of electrodes

and periods of time showing noise sensitivity much weaker than that observed in the 100- to 300-ms time-window at the  $R^2$  optimized electrodes.



In addition to the analyses performed independently at each time point (Figures 5–8), we also provide a continuous measure of integration time in the two tasks. This was achieved by measuring the time it takes to integrate 50% of noise sensitivity during the first half-second following stimulus onset (Rousselet et al., 2010). Noise sensitivity in the two tasks was normalized by the maximum absolute noise sensitivity in the identity task, defined across time frames. Then the absolute noise sensitivity in each task was integrated over time (Figure 10). At the group level, noise sensitivity integration increased sharply after 100 ms and started to differ significantly between the two tasks at 227 ms after stimulus onset. The 50% integration threshold was reached 16 ms earlier in the identity task compared to the color task. About 14% less noise sensitivity was integrated in the color task relative to the identity task. Analyses performed in each subject individually provided a somewhat different picture. In keeping with group analyses, cumulated noise sensitivity started to rise at about 100 ms in most subjects. However, onset of task effects, 50% integration times and total cumulated noise sensitivity differed markedly across subjects and from the group analyses (Figure 10).

Given the discrepancy between group and individual subject analyses, it is important to consider weak power as a potential explanation for the absence of task effect in some subjects. Indeed, lack of significant effects might be due to a real absence of effects or the presence of relatively weak effects that our statistical test might miss. Although lack of power cannot be completely ruled out, it appears that subjects with significant task effects at the  $R^2$  optimized electrode had substantial effect sizes, with maximum  $F$  cluster sums at least 1.6 times larger than the largest bootstrap  $F$  cluster sums

obtained by chance (Figure 11:  $S_1 = 2.7$ ,  $S_2 = 2.6$ ,  $S_3 = 3.2$ ,  $S_6 = 1.9$ ,  $S_7 = 2.1$ ,  $S_{10} = 2.3$ ,  $S_{11} = 2.4$ ). One subject had lower effect size than the other subjects, with a cluster sum 0.8 larger than that obtained by chance (Figure 11:  $S_4$ ). Subjects with no significant task effects had no significant cluster whatsoever ( $S_5$ ), relatively low cluster sums ( $S_9$  and  $S_{12}$ ) or cluster sums so low that they fell at the bottom of the bootstrap cluster sum distributions ( $S_8$  and  $S_{13}$ ).

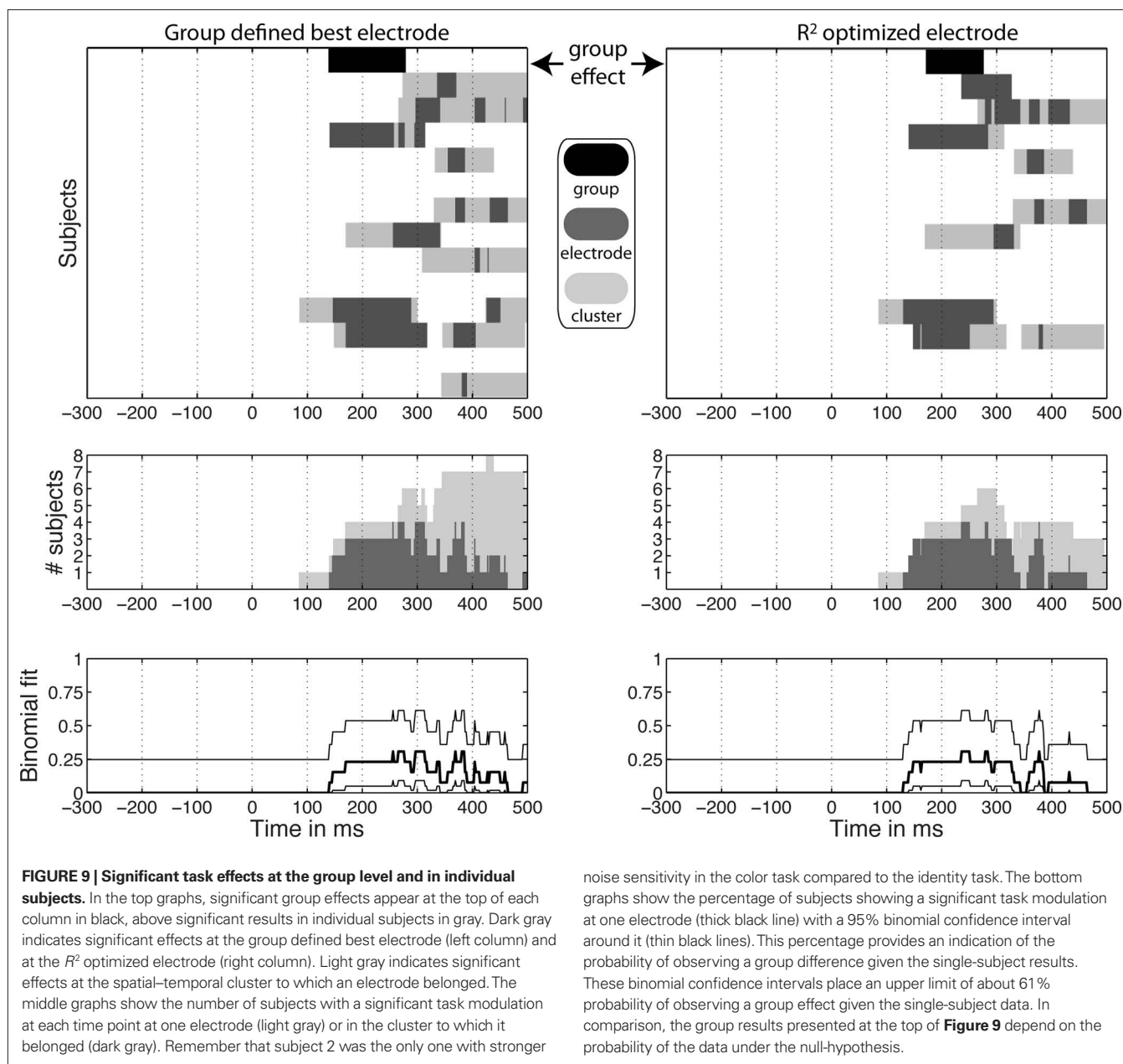
### SHIFT FUNCTION ANALYSES OF THE DECILES OF THE SINGLE-TRIAL ERP DISTRIBUTIONS

Changes in task constraints could affect noise sensitivity by modulating preferentially single-trial ERPs to noise textures or to faces. Alternatively, these changes could be a uniform compression or expansion of the distribution. In our design, noise levels are artificial. Therefore, we studied the nature of the task effects using the shift function, a technique that assumes data follow a continuum. The shift function analyses revealed that the modulation in noise sensitivity in the color task could be attributed to a modulation of a particular type of stimuli. In five subjects (Figure 12:  $S_1$ ,  $S_3$ ,  $S_7$ ,  $S_{10}$ ,  $S_{11}$ ), noise sensitivity reduction in the color task was due primarily to increased amplitudes of face ERPs, which became closer to that of noise trials. In two subjects (Figure 12:  $S_4$  and  $S_6$ ), noise sensitivity reduction was due mostly to an increase in amplitude of the noise trials. Finally, in the only subject who showed increased noise sensitivity in the color task ( $S_2$ ), the effect was also due to an amplitude increase of ERPs to noise trials. In addition, in  $S_2$ ,  $S_4$ ,  $S_6$ ,  $S_7$ , and  $S_{10}$ , there was an overall increase in ERP amplitude in the color task compared to the identity task. Thus, task constraints had non-uniform effects on ERP distributions, with most modulations being an increase in amplitude of face trials.

### DISCUSSION

Using identical colored face stimuli in two tasks, and a parametric noise manipulation, we observed a significant reduction in ERP noise sensitivity when noise level was task irrelevant. Overall, following (Philiastides et al., 2006) we conclude that task effects on noise sensitivity are weak before 200 ms, in the window of the N170, and mainly present around the P2. However, task effects were highly variable across subjects, with individual differences in onsets, durations and effect sizes. These idiosyncrasies will need to be addressed in future studies.

Based on the work of (Philiastides et al., 2006), we tested the hypothesis that there is a clear boundary, at about 200 ms after stimulus onset, between bottom-up face processing and brain activity that depends on task demand. Our group results were qualitatively similar to those of Philiastides et al. (2006) with weak task effects before 200 ms and stronger differences beyond 200 ms. Changing the task requirements did not abolish noise sensitivity altogether, but reduced it by about 19–46% in individual subjects. Results were also inconsistent across subjects, with a minority of subjects showing effects consistent with group analyses, and several subjects showing no significant effects whatsoever, despite similar behavioral performances. The nature of the task effects also differed among subjects, as revealed by analyses of single-trial ERP distributions. These results point to the existence of idiosyncratic modulations of brain activity depending on task requirements.

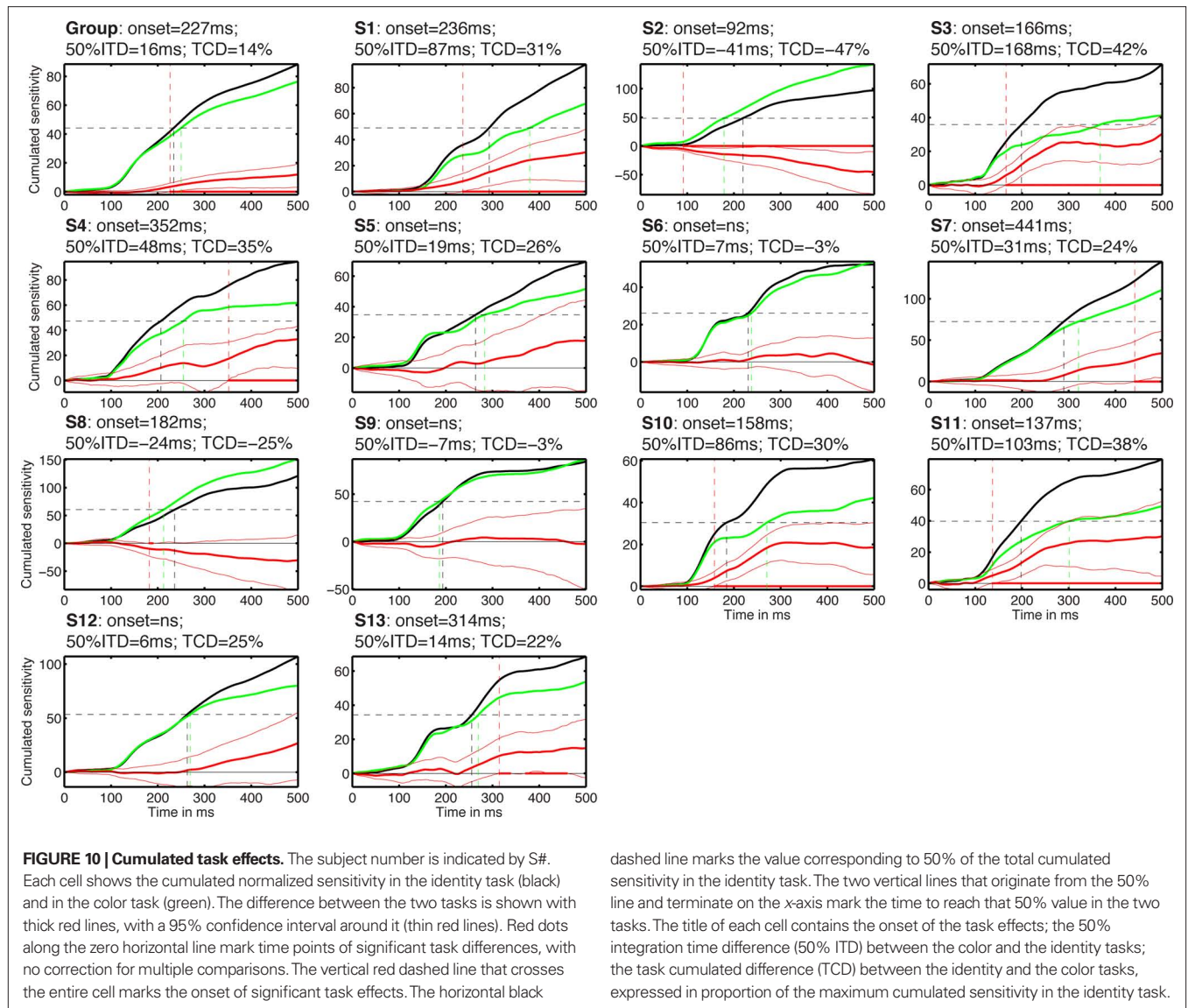


It remains unclear whether the task dependent noise sensitivity we observed is related to differences in task difficulty between the color and the identity tasks, or if it is due to changes in the diagnosticity of stimulus phase information (Banko et al., 2011; Philiastides et al., 2006; Philiastides and Sajda, 2007). More generally, noise sensitivity between 100 and 300 ms after stimulus onset probably reflects the activity of object and face processing areas that are sensitive to stimulus evidence (Philiastides and Sajda, 2007; Rousselet et al., 2008b; Tjan et al., 2006). Noise sensitivity however does not reflect activity from a general discrimination mechanism because it is not present for color and identity. Similarly, Philiastides et al. (2006) found that early single-trial visual ERPs did not discriminate between red and green or between two motion directions. However, these other

stimulus dimensions can be studied by using different techniques, such as adaptation (Vizioli et al., 2010), ICA and filtering (Snyder and Foxe, 2010), and frequency tagging (Quigley et al., 2010).

Contrary to several ERP studies described in the introduction, some of our subjects did show moderate task modulations in the time-window of the N170. The absence of task effects in previous face ERP studies is difficult to interpret because of the use of group statistics. One thing that most studies have in common is the use of relatively high-contrast stimuli. Because the effect of attention is contrast dependent (Reynolds and Heeger, 2009), attention effects on face ERPs might be more likely to be observed at low-contrast. A systematic study of attention modulations as a function of face contrast remains to be

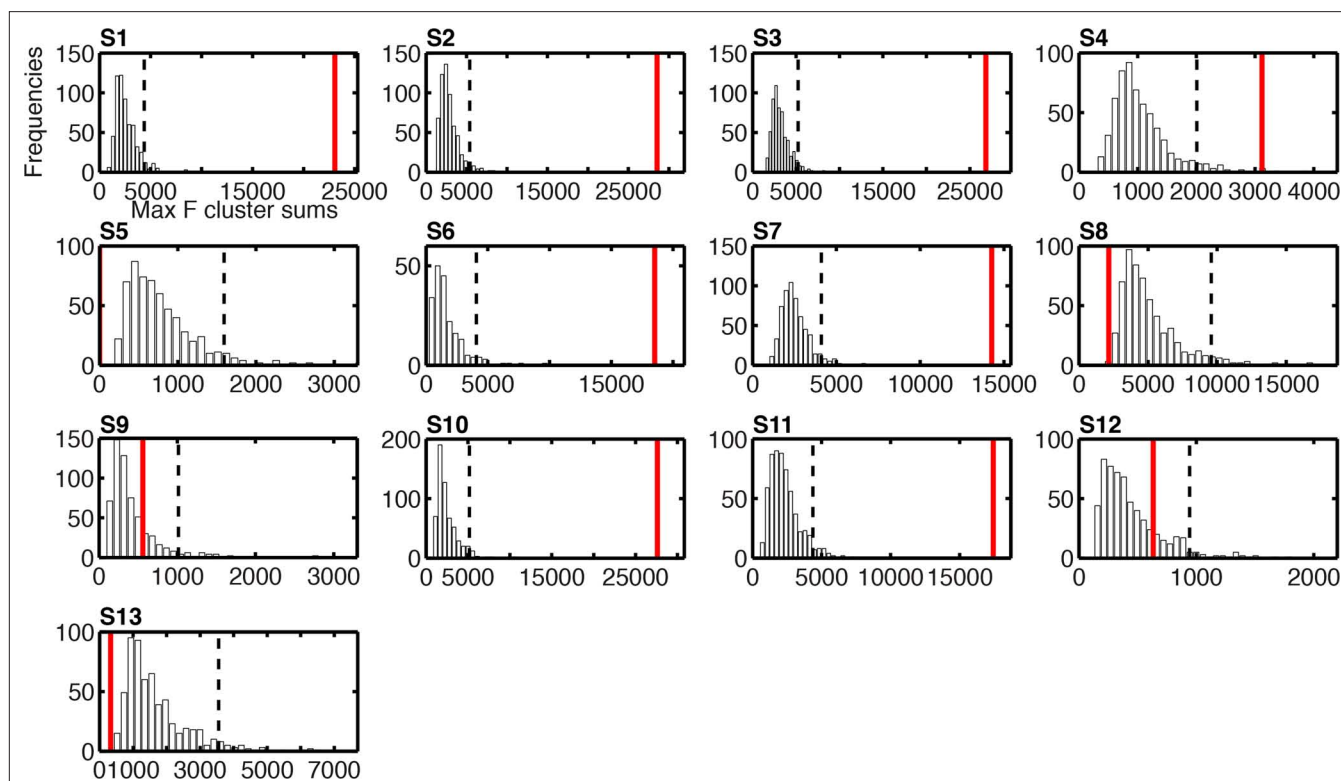




carried out. In our experiment, contrast and luminance were quite low, suggesting even weaker task modulations in more realistic circumstances. Another problem with previous reports of null effects in group statistics is the absence of statistical analyses in individual subjects, as well as poor data description (Rousselet and Pernet, 2011). Group statistics pull out effects that are consistent across subjects, even when these effects are not significant in individual subjects. Although this might seem like a good property, given the number of face ERP experiments carried out each year and the common belief that it is satisfactory to test more subjects to achieve significance, group statistics might be responsible for many false positives in the literature (Wagenmakers, 2007). By definition, group statistics are also insensitive to single-subject significant effects that are inconsistent across subjects, for instance if timing and topographies differ. Thus, group statistics can be misleading because of increased chances of false positives and false negatives, at least in theory.

Yet, it is at the moment difficult to evaluate if our results, showing a large discrepancy between group and single-trial analyses constitute a unique curiosity or if our results reveal a pervasive problem in the ERP literature. Indeed, typical face ERP studies are mostly concerned with group statistics of peak measurements, with little concern for reliability and quantification of the effects. In fact, most studies are content with the discussion of any effect  $p < 0.05$  (Rousselet and Pernet, 2011). Current practice in the ERP literature tends to hide the rich inter-subject variability that we ought to explain: we perform perceptual tasks as individuals, not as a collective brain. Moreover, many studies report weak effect sizes, unexpected results and do not control for multiple comparisons properly. One is left wondering what proportion of ERP results will ever be replicated (Miller, 2009). In many studies, beyond the recurrent fundamental flaws of null-hypothesis significance testing (Wagenmakers, 2007), the lack of robustness of  $t$ -tests and ANOVAs, and the lack of proper





**FIGURE 11 | Histograms of the bootstrap distributions of maximum  $F$  cluster sums for task effects on single-trial ERP noise sensitivity.** The subject number is indicated by S# in bold font. These bootstrap distributions were calculated under the null-hypothesis  $H_0$ , as described in Methods; hence they reflect the size of spatial-temporal task effects that can be expected by chance, due to random sampling, across the entire search space. For each

subject, the vertical black dashed line marks the 95<sup>th</sup> percentile of the  $H_0$  bootstrap distribution. The vertical red continuous line indicates the maximum sum of  $F$  values across the spatial-temporal clusters that contained the maximum  $R^2$  electrode. For subject  $S_9$ , the cluster sum is equal to zero because no cluster passed the two-electrode threshold: they were present at this electrode only.

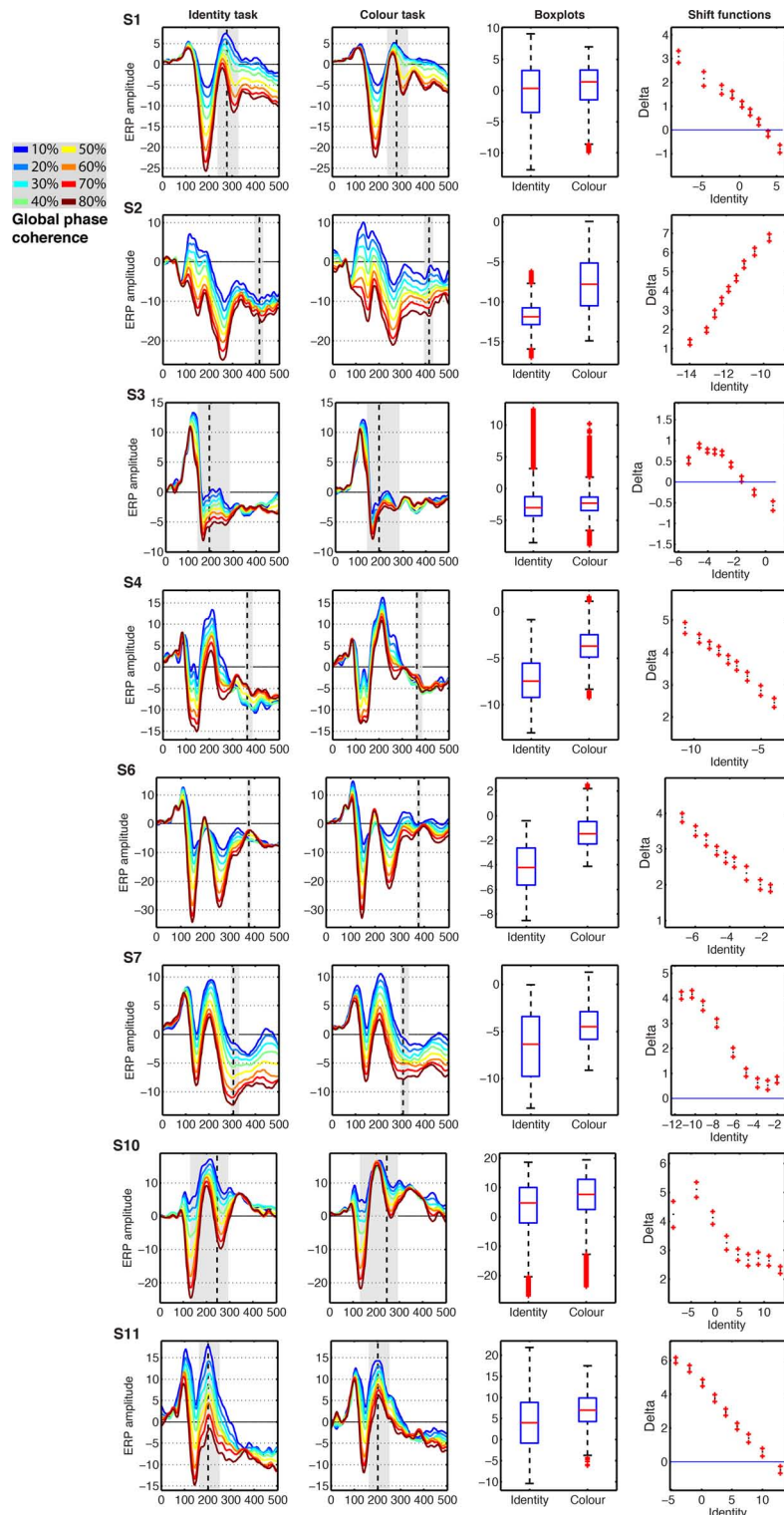
control for multiple comparisons (Wilcox, 2005), readers are too often left with so little evidence that it is impossible to judge the importance of their results. Here we've tried to provide a richer set of descriptions than is usually available in face ERP papers.

Of course, the lack of significant task effects in some subjects, and the lack of consistency across subjects who did show significant effects, might be attributable to different sources of variance, including differences in scalp thickness and electrode application, rather than individual differences in visual processing. These differences could lead to differences in statistical power across subjects. Although subjects who did show task effects had relatively large effect sizes, it is possible that more trials or better regression analyses, or both, would be necessary to reveal significant effects at different time points and in subjects showing null results. We are exploring the possibility of using smooth variance estimators, weighted models, and adjusting statistical thresholds based on empirical distributions to increase statistical power. However, our data driven estimates of effects expected by chance suggest that some subjects had indeed no task modulation of noise sensitivity whatsoever (Figure 11).

Finally, null or inconsistent effects might not reveal the absence of an effect but our failure to quantify changes in a multi-dimensional space. For instance, we reported most of our single-trial analyses at one electrode only. Based on extensive inspection

of the current dataset and previous datasets, the electrode at which the model provides the best fit seems to capture most of the effects (Rousselet et al., 2008b, 2009, 2010). Because of spatial blurring, neighboring electrodes contain redundant information, so pooling results across electrodes as is often done in group analyses would be of no benefit for univariate single-subject analyses. Of course, there might be extra information available in a multivariate space containing a large number of electrodes (Philiastides and Sajda, 2006). Hence, it will be worth extending our univariate model to measure multivariate relationships between single-trial ERP amplitude, stimulus evidence, and task demand. Finally, as discussed by (Liu et al., 2009), a single-trial linear classifier has the advantage over a GLM approach to provide a measure of information. However, it is not clear how linear classifiers can be applied to more complicated designs such as our ANCOVA.

To conclude, all these considerations about group and single-subject analyses are rather circular, because it is not clear what ought to be found. The ERP community relies mostly on group analyses, and therefore most readers might be biased to conclude that discrepancies between group and single-subject analyses reflect problems in single-subject analyses. This point of view is misguided because a significant group effect does not provide any guarantee that even 50% of the subjects will show the group effect (Figure 9).



**FIGURE 12 | Event-related potentials results from the eight subjects showing significant task effects.** The subject number is indicated by S# in bold font. The first two columns show the modeled ERPs in the identity and the color tasks. The vertical dashed line marks the latency of the largest task difference. The vertical gray shaded area marks all the continuous time frames at which a significant effect was observed, and which contained the time frame of maximum effect. The third column

shows boxplots of the single-trial modeled ERP<sub>(t<sub>est</sub>)</sub> amplitudes in the identity and color tasks, summed across the time frames marked by gray areas in columns 1 and 2. The fourth column shows the shift function between the distributions in column three. The x-axis shows the estimated deciles in the identity task. The y-axis shows the estimated difference deciles between the identity and the color task, marked as nine dots, with the ends of the confidence intervals marked by plus signs.

The inter-individual differences we describe in detail in this paper might well be differences in visual processing. If this is true, these differences must be explained, and should not be simply smoothed out by using group statistics.

## CONCLUSION

At the group level, ERP sensitivity to phase noise was reduced between about 140 and 300 ms when stimulus phase information was task irrelevant. This result suggests that sensitivity to image structure can be modulated by the task at hand, but only after an initial period of essentially bottom-up visual processing, from about 90 to 140 ms. In contrast to what the group results might suggest, we observed a significant task effect in only 60% of subjects, and at any time point only 31% of subjects showed results consistent with group analyses. Interestingly, the origin of the reduction in ERP sensitivity to phase information differed among subjects, being due to an increase in ERP amplitude to the noisiest stimuli

(noise textures) in some subjects and an increase in ERP amplitude to the least noisy stimuli (faces) in others. Overall, our analyses demonstrate the usefulness of single-trial analyses and parametric designs to study information processing. Our results also suggest that, in some situations, group statistics can be so misleading that their use, without complementary individual subject analyses, is questionable.

## ACKNOWLEDGMENTS

ESRC grant RES-000-22-3209 supported Guillaume A. Rousselet. Cyril R. Pernet is funded by the SINAPSE collaboration – <http://www.sinapse.ac.uk>, a pooling initiative funded by the Scottish Funding Council and the Chief Scientist Office of the Scottish Executive. A Biomedical Vacation Scholarship from the Wellcome Trust and a grant from the University of Glasgow Settlement charity supported Kacper P. Wiczorek. We thank Sarah Driscoll for her help collecting data.

## REFERENCES

- Aranda, C., Madrid, E., Tudela, P., and Ruz, M. (2010). Category expectations: a differential modulation of the N170 potential for faces and words. *Neuropsychologia* 48, 4038–4045.
- Banko, E. M., Gal, V., Kortvetyes, J., Kovacs, G., and Vidnyanszky, Z. (2011). Dissociating the effect of noise on sensory processing and overall decision difficulty. *J. Neurosci.* 31, 2663–2674.
- Berkovits, I., Hancock, G. R., and Nevitt, J. (2000). Bootstrap resampling approaches for repeated measure designs: relative robustness to sphericity and normality violations. *Educ. Psychol. Meas.* 60, 877–892.
- Bötzel, K., and Grüsser, O.-J. (1989). Electric brain potentials evoked by pictures of faces and non-faces: a search for “face-specific” EEG-potentials. *Exp. Brain Res.* 77, 349–360.
- Carmel, D., and Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition* 83, 1–29.
- Crist, R. E., Wu, C.-T., Karp, C., and Woldorff, M. G. (2008). Face processing is gated by visual spatial attention. *Front. Hum. Neurosci.* 1:10. doi: 10.3389/neuro.09.010.2007
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.
- Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34, 1443–1449.
- Eimer, M. (2000). Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Brain Res. Cogn. Brain Res.* 10, 145–158.
- Foxe, J. J., and Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. A framework for defining “early” visual processing. *Exp. Brain Res.* 142, 139–150.
- Furey, M. L., Tanskanen, T., Beauchamp, M. S., Avikainen, S., Uutela, K., Hari, R., and Haxby, J. V. (2006). Dissociation of face-selective cortical responses by attention. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1065–1070.
- Gaspar, C. M., Rousselet, G. A., and Pernet, C. R. (in press). Reliability of ERP and single-trial analyses. *NeuroImage*.
- Gaspar, C. M., and Rousselet, G. A. (2009). How do amplitude spectra influence rapid animal detection? *Vision Res.* 49, 3001–3012.
- Goffaux, V., Jemel, B., Jacques, C., Rossion, B., and Schyns, P. G. (2003). ERP evidence for task modulations on face perceptual processing at different spatial scales. *Cogn. Sci.* 27, 313–325.
- Gold, J., Bennett, P. J., and Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Res.* 39, 3537–3560.
- Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2007). Inverting houses and textures: investigating the characteristics of learned inversion effects. *Vision Res.* 47, 3350–3359.
- Jacques, C., and Rossion, B. (2008). Electrophysiological evidence for temporal dissociation between spatial attention and sensory competition during human face processing. *Cereb. Cortex* 17, 1055–1065.
- Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Exp. Brain Res.* 78, 193–202.
- Jeffreys, D. A., and Tuckmachi, E. S. (1992). The vertex-positive scalp potential evoked by faces and by objects. *Exp. Brain Res.* 91, 340–350.
- Jeffreys, D. A., Tuckmachi, E. S., and Rockley, G. (1992). Evoked potential evidence for human brain mechanisms that respond to single, fixated faces. *Exp. Brain Res.* 91, 351–362.
- Johnson, J. S., and Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *J. Vis.* 3, 499–512.
- Kovesi, P. (1999). Image features from phase congruency. *Videre J. Comput. Vis. Res.* 1, 2–26.
- Kovesi, P. (2003). “Phase congruency detects corners and edges,” in *Paper presented at The Australian Pattern Recognition Society Conference: DICTA 2003*, Sydney.
- Landau, A. N., Aziz-Zadeh, L., and Ivry, R. B. (2010). The influence of language on perception: listening to sentences about faces affects the perception of faces. *J. Neurosci.* 30, 15254–15261.
- Liu, H., Agam, Y., Madsen, J. R., and Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62, 281–290.
- Liu, J., Harris, A., and Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nat. Neurosci.* 5, 910–916.
- Lueschow, A., Sander, T., Boehm, S. G., Nolte, G., Trahms, L., and Curio, G. (2004). Looking for faces: attention modulates early occipitotemporal object processing. *Psychophysiology* 41, 350–360.
- Makeig, S., Delorme, A., Westerfield, M., Jung, T. P., Townsend, J., Courchesne, E., and et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol.* 2, e176. doi: 10.1371/journal.pbio.0020176
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190.
- Miller, J. (2009). What is the probability of replicating a statistically significant effect? *Psychon. Bull. Rev.* 16, 617–640.
- Mohamed, T. N., Neumann, M. F., and Schweinberger, S. R. (2009). Perceptual load manipulation reveals sensitivity of the face-selective N170 to attention. *Neuroreport* 20, 782–787.
- Nobre, A. C., Allison, T., and McCarthy, G. (1998). Modulation of human extrastriate visual processing by selective attention to colours and words. *Brain* 121, 1357–1368.
- Okazaki, Y., Abrahamyan, A., Stevens, C. J., and Ioannides, A. A. (2008). The timing of face selectivity and attentional modulation in visual processing. *Neuroscience* 152, 1130–1144.
- Pernet, C., Schyns, P. G., and Demonet, J. F. (2007). Specific, selective or preferential: comments on category specificity in neuroimaging. *Neuroimage* 35, 991–997.
- Pernet, C. R., Chauveau, N., Gaspar, C., and Rousselet, G. A. (2011). LIMO EEG: a toolbox for hierarchical linear MOdeling of ELeTroEncephaloGraphic data. *Comput. Intell. Neurosci.* [Article ID 831409]. doi: 10.1155/2011/831409
- Philastides, M. G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *J. Neurosci.* 26, 8965–8975.
- Philastides, M. G., and Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cereb. Cortex* 16, 509–518.

- Philiastides, M. G., and Sajda, P. (2007). EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *J. Neurosci.* 27, 13082–13091.
- Puce, A., Allison, T., and McCarthy, G. (1999). Electrophysiological studies of human face perception. III: effects of top-down processing on face-specific potentials. *Cereb. Cortex* 9, 445–458.
- Quigley, C., Andersen, S. K., Schulze, L., Grunwald, M., and Müller, M. M. (2010). Feature-selective attention: evidence for a decline in old age. *Neurosci. Lett.* 474, 5–8.
- Ratcliff, R., Philiastides, M. G., and Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6539–6544.
- Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185.
- Rossion, B., and Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage* 39, 1959–1979.
- Rousselet, G. A., Gaspar, C. M., Pernet, C. R., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2010). Healthy aging delays scalp EEG sensitivity to noise in a face discrimination task. *Front. Psychology* 1:19. doi: 10.3389/fpsyg.2010.00019
- Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2007a). Single-trial EEG dynamics of object and face visual processing. *Neuroimage* 36, 843–862.
- Rousselet, G. A., Macé, M. J.-M., Thorpe, S. J., and Fabre-Thorpe, M. (2007b). Limits of ERP differences in tracking object processing speed. *J. Cogn. Neurosci.* 19, 1–18.
- Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2008a). Time course and robustness of ERP object and face differences. *J. Vis.* 8:3, 1–18.
- Rousselet, G. A., Pernet, C. R., Bennett, P. J., and Sekuler, A. B. (2008b). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience* 9, 98. doi: 10.1186/1471-2202-9-98
- Rousselet, G. A., Husk, J. S., Pernet, C. R., Gaspar, C. M., Bennett, P. J., and Sekuler, A. B. (2009). Age-related delay in information accrual for faces: evidence from a parametric, single-trial EEG approach. *BMC Neurosci.* 10, 114. doi: 10.1186/1471-2202-10-114
- Rousselet, G. A., and Pernet, C. R. (2011). Quantifying the time course of visual object processing using ERPs: it's time to up the game. *Front. Psychology* 2:107. doi: 10.3389/fpsyg.2011.00107
- Rousselet, G. A., Thorpe, S. J., and Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends Cogn. Sci.* 8, 363–370.
- Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition* 67, 147–179.
- Schyns, P. G. (2010). Grand challenges in perception science: modeling the future. *Front. Psychology* 1:10. doi: 10.3389/fpsyg.2010.00010
- Schyns, P. G., Gosselin, F., and Smith, M. L. (2009). Information processing algorithms in the brain. *Trends Cogn. Sci. (Regul. Ed.)* 13, 20–26.
- Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., and Gosselin, F. (2003). A principled method for determining the functionality of brain responses. *Neuroreport* 14, 1665–1669.
- Schyns, P. G., Petro, L. S., and Smith, M. L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. *Curr. Biol.* 17, 1580–1585.
- Seco, G. V., Izquierdo, M. C., García, M. P. E., and Díez, F. J. H. (2006). A comparison of the Bootstrap-F, improved general approximation, and Brown-Forsythe multivariate approaches in a mixed repeated measures design. *Educ. Psychol. Meas.* 66, 35–62.
- Seeck, M., and Grüsser, O.-J. (1992). Category-related components in visual evoked potentials: photographs of faces, persons, flowers and tools as stimuli. *Brain Res.* 92, 338–349.
- Séverac-Cauquil, A., Edmonds, G. E., and Taylor, M. J. (2000). Is the face-sensitive N170 the only ERP not affected by selective attention? *Neuroreport* 11, 2167–2171.
- Smith, M. L., Gosselin, F., and Schyns, P. G. (2007). From a face to its category via a few information processing states in the brain. *Neuroimage* 37, 974–984.
- Snyder, A. C., and Foxe, J. J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. *J. Neurosci.* 30, 4024–4032.
- Sreenivasan, K. K., Katz, J., and Jha, A. P. (2007). Temporal characteristics of top-down modulations during working memory maintenance: an event-related potential study of the N170 component. *J. Cogn. Neurosci.* 19, 1836–1844.
- Thorpe, S. J. (2009). The speed of categorization in the human visual system. *Neuron* 62, 168–170.
- Thorpe, S. J., and Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science* 291, 260–263.
- Tjan, B. S., Lestou, V., and Kourtzi, Z. (2006). Uncertainty and invariance in the human visual cortex. *J. Neurophysiol.* 96, 1556–1568.
- van Rijsbergen, N. J., and Schyns, P. G. (2009). Dynamics of trimming the content of face representations for categorization in the brain. *PLoS Comput. Biol.* 5, e1000561. doi: 10.1371/journal.pcbi.1000561
- Vizioli, L., Rousselet, G. A., and Caldara, R. (2010). Neural repetition suppression to identity is abolished by other-race faces. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20081–20086.
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804.
- Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing*, 2nd Edn. New York, NY: Elsevier Academic Press.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 January 2011; accepted: 09 June 2011; published online: 23 June 2011.  
Citation: Rousselet GA, Gaspar CM, Wiczorek KP and Pernet CR (2011) Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects). *Front. Psychology* 2:137. doi: 10.3389/fpsyg.2011.00137  
This article was submitted to *Frontiers in Perception Science*, a specialty of *Frontiers in Psychology*.  
Copyright © 2011 Rousselet, Gaspar, Wiczorek and Pernet. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.